

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/59126 A2

(51) International Patent Classification⁷: C12N 15/52,
C12P 17/18, 19/62, C07H 17/08

(21) International Application Number: PCT/GB01/00509

(22) International Filing Date: 8 February 2001 (08.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0002840.7 8 February 2000 (08.02.2000) GB
0008786.6 10 April 2000 (10.04.2000) GB
0009387.2 14 April 2000 (14.04.2000) GB

(71) Applicants (for all designated States except US):
NORGES TEKNISK NATURVITENSKAPELIGE
UNIVERSITET [NO/NO]; N-7491 Trondheim
(NO). STIFTELSEN FOR INDUSTRIELL OG
TEKNISK FORSKNING VED NORGES TEKNISKE
HØGSKOLE [NO/NO]; Sintef Applied Chemistry Di-
vision, Sem Seelands Vel 2A, N-7034 Trondheim (NO).

ALPHARMA AS [NO/NO]; Harbitzalléen 3, N-0212
Oslo (NO). SINVENT AS [NO/NO]; Strindveien 4,
N-7465 Trondheim (NO).

(71) Applicant (for GB only): DZIEGLEWSKA, Hanna
[GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria
Street, London EC4V 4EL (GB).

(71) Applicants and

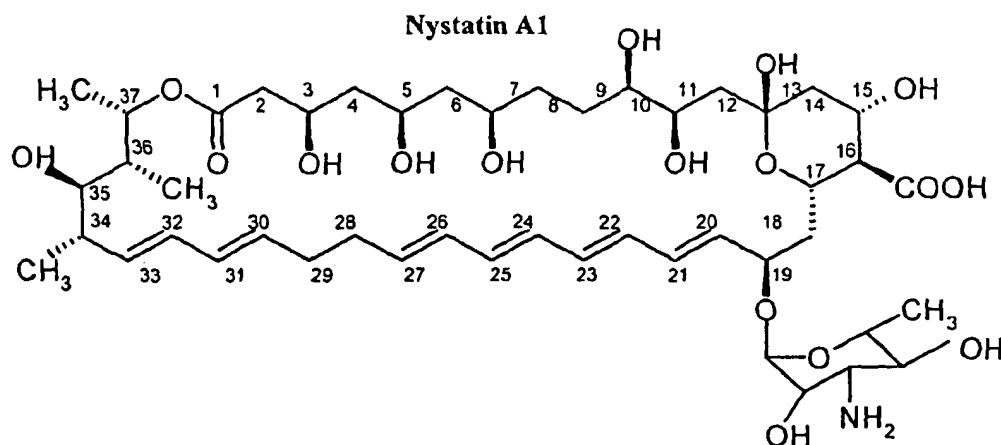
(72) Inventors: ZOTCHEV, Sergey Borisovich [RU/NO];
Elgester gate 51, N-7030 Trondheim (NO). SEKUROVA,
Olga Nikolayivna [RU/NO]; Elgester gate 51, N-7030
Trondheim (NO). FJÆRVIK, Espen [NO/NO]; Lianveien
10, N-7025 Trondheim (NO). BRAUTASET, Trygve
[NO/NO]; Martin Stokkens vei 74, N-7027 Trondheim
(NO). STRØM, Arne, Reidar [NO/NO]; Konvallvegen
38, N-7050 Trondheim (NO). VALLA, Svein [NO/NO];
Naustanveien 2, N-7560 Vikhamar (NO).

(72) Inventors; and

(75) Inventors/Applicants (for US only): ELLINGSEN,
Trond, Erling [NO/NO]; Markaplassen 273, N-7054
Ranheim (NO). SLETTA, Håvard [NO/NO]; Uglaveien

[Continued on next page]

(54) Title: NOVEL GENES ENCODING A NYSTATIN POLYKETIDE SYNTHIASE AND THEIR MANIPULATION AND
UTILITY



(57) Abstract: The invention provides a nucleic acid molecule comprising: (a) a nucleotide sequence as shown in SEQ ID No. 35; or (b) a nucleotide sequence which is the complement of SEQ ID No. 35; or (c) a nucleotide sequence which is degenerate with SEQ ID No. 35; or (d) a nucleotide sequence hybridising under conditions of high stringency to SEQ ID No. 35, to the complement of SEQ ID No. 35, or to a hybridisation probe derived from SEQ ID No. 35 or the complement thereof; or (e) a nucleotide sequence having at least 80 % sequence identity with SEQ ID No. 35; or (f) a nucleotide sequence having at least 65 % sequence identity with SEQ ID No. 35 wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof. Also provided are part of such molecules and polypeptides (and parts thereof) encoded by such a nucleic acid molecule, and the use of such molecules and polypeptides in facilitating nystatin biosynthesis and in the synthesis of nystatin derivatives and novel polyketide as macrolide structures.



19, N-7024 Trondheim (NO). GULLIKSEN, Ole-Martin [NO/NO]; Grevlingvn. 31 D, N-0595 Oslo (NO).

(utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(74) Agents: DZIEGLEWSKA, Hanna et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(81) Designated States (national): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

GENE CLUSTER ENCODING A NYSTATIN POLYKETIDE
SYNTHASE AND ITS MANIPULATION AND UTILITY

5

The present invention relates to the cloning and sequencing of the gene cluster encoding a modular polyketide synthase enzyme involved in the biosynthesis of the macrolide antibiotic nystatin. The invention thus relates to novel genes and nucleic acid molecules encoding proteins/polypeptides exhibiting functional activities involved in nystatin biosynthesis, such functional proteins/polypeptides themselves, and their uses both in facilitating nystatin biosynthesis and in the synthesis of nystatin derivatives and novel polyketide or macrolide structures.

Polyketides are natural products synthesized by microorganisms, many of which have applied potential as pharmaceuticals or as agricultural or veterinary products. Examples of polyketides used in medical treatments include the antibiotics erythromycin (antibacterial), nystatin (antifungal), avermectin (antiparasitic), rapamycin (immunosuppressant) and daunorubicin (antitumor). The Gram-positive bacteria *Streptomyces* are the main producers of polyketides, and the genetics and biochemistry of polyketide biosynthesis in these organisms are relatively well characterized (Hopwood et al., Chem. Rev. v. 97: 2465-2497 (1997)). Macrolide polyketide compounds are formed via repeated condensations of simple carboxylic acids by modular (type I) polyketide synthases (PKS) in a manner similar to fatty acid biosynthesis. The modular hypothesis proposed by Donadio et al. Science, v. 252: 675-679 (1991) suggested that type I PKSs are organized in repeated units (modules), each of which is responsible for one condensation cycle in the synthesis of a polyketide chain. This was proven to be correct by

- 2 -

manipulations of PKSs type I genes resulting in predictable changes in the chemical structures of macrolides. Beside condensation of the next carboxylic acid onto the growing polyketide chain, ensured by the catalytic activity of the β -ketoacyl synthase (KS) domain, modules of PKSs type I may contain domains with β -ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) activities, which determine the reduced state of incorporated extender units. The acyltransferase (AT) and acyl carrier protein (ACP) domains present in each module are responsible for the choice of extender unit and retention of the growing polyketide chain on the PSK, respectively. Upon completion of synthesis, the polyketide chain is released from PKSs via action of a thioesterase (TE), that is probably also involved in cyclization of the final product. Thus, PKSs type I represent an assembly line for polyketide biosynthesis, that can be manipulated by changing the number of modules, their specificities towards carboxylic acids, and by inactivating or inserting domains with reductive activities (Katz, Chem. Rev., v. 97, 2557-2575, 1997). After the polyketide moiety of a macrolide is synthesized and cyclized to form a macrolactone ring, it is usually modified via hydroxylation, glycosylation, methylation and/or acylation. These modifications are believed to be crucially important for the biological activities of macrolides.

The genes for macrolide antibiotic biosynthesis in *Streptomyces* are organized in clusters, and a number of such clusters have already been identified. Exploitation of recombinant DNA technology has made it possible to isolate complete antibiotic biosynthetic gene clusters by screening the gene libraries with DNA probes encoding PKSs (Schwecke et al., Proc. Natl. Acad. Sci. USA, v. 92: 7839-7843, 1995). The molecular cloning and complete DNA sequencing has been described

- 3 -

for several macrolide antibiotic gene clusters of *Streptomyces*, including those for avermectin, pikromycin and rapamycin (Ikeda et al., Proc. Natl. Acad. Sci. USA, v. 96: 9509-9514, 1999; Xue et al., Proc. Natl. Acad. Sci. USA, v. 95: 1211-12116, 1998; Schwecke et al., Proc. Natl. Acad. Sci. USA, v. 92: 7839-7843, 1995). Partial cloning and DNA sequencing of the gene cluster for the polyene macrolide antibiotic pimarin has recently been reported (Aparicio et al., J. Biol. Chem., v. 274: 10133-10139, 1999). However, a complete DNA sequence of genes for the biosynthesis of a polyene macrolide antibiotic with antifungal activity has not yet been disclosed. There is a need and desire to increase the repertoire of available antifungal antibiotics, and/or to improve upon the properties (e.g. efficacy, toxicity, etc.) of existing drugs. Hence the provision of new antifungal treatments, particularly those exhibiting new or improved properties would represent a considerable advance in the art.

The present invention is directed to this aim, and is based on the cloning and DNA sequencing of the nystatin biosynthesis gene cluster. This provides the first example of the identification of such antifungal antibiotic biosynthesis genes, as well as a tool for genetic manipulation in order to modify the properties of nystatin and/or the producing organism, or to obtain novel potentially useful compounds.

The polyene antifungal antibiotic nystatin A1, the complete stereostructure of which (see Fig. 1) has been determined by Lancelin & Beau, Tetrahedron Lett, v. 30: 4521-4524, (1989), is produced by *Streptomyces noursei* ATCC11455. From the structure of the nystatin molecule, which belongs to the class of macrolide compounds, we predicted that its polyketide backbone is synthesized by a PKS type I enzyme. Based on this assumption, and as described in the Examples below, a genomic library of *Streptomyces noursei* ATCC11455 was screened using a

- 4 -

5 specially designed probe obtained by PCR using primers based on conserved amino acid sequences within known β -ketoacyl synthase (KS) and acyl carrier protein (ACP) domains of known modular PKS enzymes. This led to the identification of a number of clones or fragments which we have sequenced and shown to contain parts or portions of the nystatin PKS gene cluster. We have further shown that alteration of the fragment sequences to inactivate the encoded product leads to abrogation of nystatin biosynthesis (see Example 1 below), thereby confirming the requirement of the identified PKS for nystatin biosynthesis. Subsequent work on the clones/fragments has lead to the sequencing of the PKS gene cluster, and the identification of the different modules and enzymatic domains, regulatory regions etc, within it.

10 Furthermore, as will be described in more detail below, we have shown that manipulations of functional DNA sequences within the novel nystatin PKS gene cluster which we have identified, have led to the synthesis of novel molecular structures, e.g. nystatin derivatives with improved function. This opens up the exciting possibility of manipulating the nystatin A1 PKS gene cluster to obtain not only beneficial new nystatin derivatives, but also to improve and facilitate the biosynthetic production process (for example to improve yield, or production conditions, or to expand the range of available host cells) or to provide novel compounds with new activities and/or properties.

25 More particularly, two primary regions (or "parts" or "portions") of the nystatin PKS gene cluster were initially identified and sequenced, together representing approximately 80% of the nystatin PKS gene cluster, and the functional sequences within said regions (e.g. PKS genes, regulatory regions etc, as well as functional gene products, enzymatic domains etc.) have been identified and characterised.

The first region ("Region 1"), which we have termed

- 5 -

"Nys 1" and the complete DNA sequence of which is shown in SEQ ID No. 1, has been shown to contain a number of PKS or associated genes or regulatory regions, and 13 separate "features" or open reading frames (ORFs) have been identified (the amino acid sequences of the translation products of which are shown in SEQ ID Nos. 3 to 15 respectively).

The second region ("Region 2"), which we have termed "Nys 2", and the complete coding sequence of which is shown in SEQ ID No. 2, also comprises a number of "functional" regions, and 5 separate "features" or ORFs have been identified, the amino acid sequences of the translation products of which are shown in SEQ ID Nos. 16 to 20 respectively.

Nys 1 and Nys 2 (i.e. SEQ ID Nos. 1 and 2 and the sequences they encode) are the subject of British Patent Application No. 0002840.7 filed on 8 February 2000.

Subsequent sequencing efforts have led to the determination of the sequence of the DNA spanning the gap between SEQ ID Nos. 1 and 2, and the identification of novel genes in this region. In addition, the partial gene sequences contained in SEQ ID Nos. 1 and 2 encoding the gene products NysI (SEQ ID No. 20) and NysDII (SEQ ID Nos. 3) (see further below) have been completed (see new SEQ ID Nos. 36 and 37 respectively - see further below). Thus, these sequencing efforts have led to the identification and sequencing of the DNA region encompassing the entire nystatin PKS gene cluster, and the identification and characterisation of the functional sequences within this region.

The complete coding sequence for (i.e. the complete nucleotide sequence encoding) the nystatin biosynthetic gene cluster is shown in SEQ ID No. 35. This has been shown to contain a number of PKS or associated genes or regulatory regions, and 23 separate "features" or ORFs have been identified (the amino acid sequences of the translation products of which are shown in SEQ ID Nos. 2

- 6 -

to 19, and 36 to 42 respectively).

The complete coding sequence for (i.e. the complete nucleotide sequence encoding) the nystatin biosynthetic gene cluster, as shown in SEQ ID No. 35, is the subject
5 of British Patent Application No. 0008786.6 filed on 10 April 2000 and British Patent Application No. 0009387.2 filed on 14 April 2000.

In one aspect, the present invention thus provides a nucleic acid molecule comprising:

- 10 (a) a nucleotide sequence as shown in SEQ ID No. 35; or
- (b) a nucleotide sequence which is the complement of SEQ ID No. 35; or
- (c) a nucleotide sequence which is degenerate with SEQ ID No. 35; or
- 15 (d) a nucleotide sequence hybridising under conditions of high stringency to SEQ ID No. 35, to the complement of SEQ ID No. 35, or to a hybridisation probe derived from SEQ ID No. 35 or the complement thereof; or
- (e) a nucleotide sequence having at least 80% sequence
20 identity with SEQ ID No. 35; or
- (f) a nucleotide sequence having at least 65% sequence identity with SEQ ID No. 35 wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof.

25 A "nystatin PKS enzyme" is defined further below, but briefly in the context of section (f) above means an enzyme or protein or polypeptide that is functional in the synthesis, transport or transfer of a macrolide antibiotic or polyketide moiety, preferably nystatin or
30 a nystatin derivative or nystatin-related molecule.

In a further aspect, the present invention also provides a nucleic acid molecule comprising:

- (a) a nucleotide sequence as shown in SEQ ID No. 1 and/or in SEQ ID No. 2; or
- 35 (b) a nucleotide sequence which is the complement of SEQ ID No. 1 and/or SEQ ID No. 2; or
- (c) a nucleotide sequence which is degenerate with SEQ

- 7 -

ID No. 1 and/or SEQ ID No. 2; or

(d) a nucleotide sequence hybridising under conditions of high stringency to SEQ ID No. 1 and/or SEQ ID No. 2, to the complement of SEQ ID No. 1 and/or SEQ ID No. 2, or to a hybridisation probe derived from SEQ ID Nos. 1 and/or 2 or the complements thereof; or

(e) a nucleotide sequence having at least 65% sequence identity with SEQ ID No. 1 and/or SEQ ID No. 2, wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof.

A nucleic acid molecule of the invention may be an isolated nucleic acid molecule (in other words isolated or separated from the components with which it is normally found in nature) or it may be a recombinant or a synthetic nucleic acid molecule.

The nucleic acid molecule of the invention encodes (or comprises a nucleotide sequence encoding) the nystatin A1 PKS enzyme, or a portion thereof e.g. a sequence encoding a single domain, or comprises a nucleotide sequence in the nystatin A1 PKS gene cluster which is a functional or non-functional genetic element. More precisely, the nucleic acid molecule of the invention encodes one or more polypeptides, or comprises one or more genetic elements having functional activity in the synthesis of a macrolide antibiotic or a polyketide moiety, preferably nystatin or a nystatin derivative or nystatin-related molecule. Such functional activity may be enzymatic activity e.g. an activity involved in the synthesis or transport or transfer of a polyketide moiety or a macrolide molecule (this can be macrolide chain or ring synthesis or any step contributory thereto, or macrolide ring or polyketide chain modification etc) and/or it may be a regulatory activity, e.g. regulation of the expression of the genes or proteins involved in the synthesis, or regulation of the synthetic process, and/or it may be a

- 8 -

"transporter activity". Thus, included generally are also transport proteins involved in the transfer or transport of polyketide or macrolide moieties e.g. in the transport or efflux of the synthesised molecule within or out of the cell. Also included in this respect are glycosylation proteins which includes molecules involved in the biosynthesis and/or attachment of saccharides (e.g. mycosamine) to the macrolide or polyketide.

Whilst nucleotide sequences encoding a desired product are preferred according to the invention, also encompassed are nucleotide sequences comprising functional genetic elements such as promoters, promoter-operator regions, enhancers, other regulatory sequences etc. Thus, the nucleic acid molecule of the invention need not comprise the entire PKS gene cluster but may comprise a portion or part of it e.g. a part encoding a polypeptide having a particular function or a regulatory sequence. This may comprise one or more genes, and/or regulatory sequences, and/or one or more modules or, enzymatic domains, or non-coding or coding functional genetic elements (e.g. elements controlling gene expression, transcription, translation etc).

In one such aspect, the invention provides a nucleic acid molecule as defined above, wherein said nucleotide sequence of SEQ ID No. 35 (or variant thereof as defined in (b) to (f) above) does not include the portions of the molecule comprising ORF 1 (see Table 1 below). In other words, in this embodiment, the nucleotide sequence of SEQ ID No. 35 does not comprise nucleotides 124026 to 125222.

In another such aspect, the invention provides a nucleic acid molecule as defined above, wherein said nucleotide sequence of SEQ ID No. 35 (or variant thereof as defined in (b) to (f) above) does not include the portions of the molecule comprising ORF 2 (see Table 1 below). In other words, in this embodiment, the

- 9 -

nucleotide sequence of SEQ ID No. 35 does not comprise nucleotides 122812 to 123876.

In a further such aspect, the invention provides a nucleic acid molecule as defined above, wherein said
5 nucleotide sequence of SEQ ID No. 35 (or variant thereof as defined in (b) to (f) above) does not include the portions of the molecule comprising *NysF* (see Table 1 below). In other words, in this embodiment, the nucleotide sequence of SEQ ID No. 35 does not comprise
10 nucleotides 454 to 1191.

Alternatively, the invention provides nucleic acid molecules which contain a part of ORF 1, ORF 2 and/or *NysF*, or a modified sequence of ORF 1, ORF 2 and/or *NysF*, such that the expression or the function of the
15 ORF 1, ORF 2 and/or *NysF* gene product is ablated.

Included within the scope of the invention are nucleotide sequences which hybridise to SEQ ID Nos. 1 or 2 or 35 or their complements, or to parts thereof (i.e. to hybridisation probes derived from SEQ ID Nos. 1 or 2
20 or 35 which are discussed in more detail below), under high stringency conditions and which preferably encode or are complementary to a sequence which encodes a nystatin PKS enzyme or part thereof. Conditions of high stringency may readily be determined according to
25 techniques well known in the art, as described for example in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2nd Edition. Hybridising sequences included within the scope of the invention are those binding under non-stringent conditions (6 x SSC/50%
30 formamide at room temperature) and washed under conditions of high stringency (e.g. 0.1 x SSC, 68°C), where SSC = 0.15 M NaCl, 0.015M sodium citrate, pH 7.2.

A hybridisation probe may be a part of the SEQ ID No. 1 or SEQ ID No. 2 or SEQ ID No. 35 sequence (or
35 complementary sequence), which is of sufficient base length and composition to function to hybridise to sample or test nucleic acid sequences to determine

- 10 -

whether or not hybridisation under high stringency condition occurs. The probe may thus be at least 15 bases in length preferably at least 30, 40, 50, 75, 100 or 200 bases in length. Representative probe lengths
5 thus include 30-500 bases e.g. 30-300, 50-200, 50-150, 75-100.

The hybridisation probe may be derived from a coding or non-coding, functional or non-functional part of the sequence (i.e. SEQ ID Nos. 1 or 2 or 35 or their
10 complements), and may for example correspond to a gene or module or to an enzymatic domain, or a part thereof (e.g. the part encoding the active site) or to a sequence which links enzymatic domains or modules. Thus, the hybridisation probe may have functional
15 activity in polyketide/macrolide synthesis as defined above.

Nucleotide sequence identity may be determined using the BestFit program of the Genetics Computer Group (GCG) Version 10 Software package from the University of
20 Wisconsin. The program uses the local homology algorithm of Smith and Waterman with the default values: Gap creation penalty = 50, Gap extension penalty = 3, Average match = 10,000, Average mismatch = -9.000.

Nucleotide sequences according to the invention may
25 exhibit at least 65%, 70%, 75%, 80%, 85%, 90%, 95% or 98% sequence identity with SEQ ID Nos. 1 or 2 or 35 and preferably encode or are complementary to a sequence which encodes a nystatin PKS enzyme or part thereof. Nucleotide sequences meeting the % sequence identity
30 criteria defined herein may be regarded as "substantially identical" sequences.

Where the nucleic acid molecules of the invention are defined by reference to SEQ ID Nos. 1 and/or 2, the nucleic acid molecule of the invention may thus comprise
35 a SEQ ID No. 1 or "SEQ ID No. 1-variant" sequence (i.e. a sequence complementary, or degenerate to SEQ ID No. 1 or a functionally equivalent variant such as a

- 11 -

hybridising or substantially identical sequence as defined above) or a SEQ ID No. 2 or "SEQ ID NO. 2-variant" sequence or both. Nucleic acid molecules comprising both a SEQ ID No. 1/SEQ ID NO. 1-variant
5 sequence and a SEQ ID No. 2/SEQ ID NO. 2-variant sequence are preferred.

As referred to herein "functionally equivalent variants" or "functional equivalents" retain at least one function of the entity to which they are related (or
10 from which they are derived), e.g. encode a protein with substantially the same properties, or exhibit substantially the same regulatory or other functional properties or activities.

As mentioned above, nucleic acid molecules
15 comprising parts or portions (e.g. fragments) of the nucleotide sequences of SEQ ID No. 35 or of SEQ ID No. 1 and/or 2 (or their complementary, degenerate or functionally equivalent variants) are also included within the scope of the invention.

20 Such parts or portions of the PKS gene cluster advantageously may also be regarded as functional equivalents of the complete sequence. For example, the sequence portion or fragment may retain a functional activity as defined above, e.g. an enzymatic,
25 regulatory, or transporter activity in polyketide or macrolide biosynthesis.

Conveniently, the part or portion of the PKS gene cluster (e.g. of SEQ ID No. 35, or 1 and/or 2) is at least 15 bases in length, more preferably at least 20,
30 25, 30, 35, 40, 50, 70, 100, 200, 300, 400, 500, 1000, 2000, 5000, 10,000, 15,000, 20,000, 30,000, or 50,000 bases. Representative fragment lengths thus include 15-50,000 bases e.g. 50-30,000 bases, or 100-20,000, 100-10,000 or 200-5,000, or 200-2,000. The part or portion
35 may comprise or encode contiguous or non-contiguous nucleotide or amino acids.

Parts or portions of functional parts of the PKS

- 12 -

gene sequences are discussed in more detail below.

Parts or portions of the PKS gene cluster may also comprise the non-coding or non-functional part of the DNA molecule (or the nucleotide sequences), for example
 5 promoter or operator sequences, or linker sequences joining individual genes, modules or enzymatic domains. These may be contiguous or non-contiguous and will be discussed further below.

As mentioned above, a number of genes and ORFs
 10 within SEQ ID Nos. 35, 1 and 2 have been identified and such genes (or their complementary, degenerate or functionally equivalent variants as defined above) represent preferred "parts" or fragments of SEQ ID Nos. 35, 1 and 2. These are tabulated in Table 1 below:

15

Table 1

Molecule features of SEQ ID Nos. 35, 1 and 2 (the whole
 gene cluster sequence (125401 bp) Nys 1 (65140 bp) and
 20 Nys 2 (27541bp) respectively)

SEQ ID No. 35				
	Start	End	Gene	Description
25	1191	454 C	<i>nysF</i>	putative 4'-phosphopantetheine transferase
	3092	1275 C	<i>nysG</i>	ABC transporter
30	4824	3070 C	<i>nysH</i>	ABC transporter
	5122	6156	<i>nysDIII</i>	dGDP-mannose-4,6-dehydratase homolog
	6338	34771	<i>nysI</i>	NysI PKS, modules 9-14
	34792	51097	<i>nysJ</i>	NysJ PKS, modules 15-17
35	51155	57355	<i>nysK</i>	NysK PKS (module 18 + TE)
	57503	58685	<i>nysL</i>	P450 monooxygenase NysL
	58980	58788 C	<i>nysM</i>	ferredoxin NysM

- 13 -

	60241	59047 C	<i>nysN</i>	P450 monooxygenase NysN
	61296	60240 C	<i>nysDII</i>	putative aminotransferase
	62837	61317 C	<i>nysDI</i>	putative UDP-glucuronosyltransferase
5	63067	67167	<i>nysA</i>	NysA PKS, loading module
	67213	76791	<i>nysB</i>	NysB PKS, modules 1 and 2
	76811	110101	<i>nysC</i>	NysC PKS, modules 3-8
	110521	111276	<i>nysE</i>	putative thioesterase
	111666	114566	<i>nysRI</i>	transcriptional activator
10	114590	117451	<i>nysRII</i>	putative transcriptional activator
	117441	120224	<i>nysRIII</i>	putative transcriptional activator
	120676	121308	<i>nysRIV</i> (short)	putative response regulator
15	120628	121308	<i>nysRIV</i> (long)	putative response regulator
	121997	122758	<i>nysRV</i>	putative repressor
	123876	122812 C	ORF2	putative transcriptional regulator
20	124026	125222	ORF1	putative peptidase (aminohydrolase)

Note: "C" indicates that the gene is encoded by the complement DNA strand.

nys 1 (SEQ ID No. 1)

	Start	End	Gene	Description Name
30	1035	3 C	<i>nysD2</i>	putative aminotransferase
	2447	1058 C	<i>nysD1</i>	putative UDP-glucuronosyl-transferase
35	2806	6904	<i>nysA</i>	nystatin PKS, loading module
	6952	16528	<i>nysB</i>	nystatin PKS, modules 1 and 2

- 14 -

	16550	49838	<i>nysC</i>	nystatin PKS, modules 3-8
	50227	51013	<i>nysE</i>	putative thioesterase
	51405	54303	<i>nysR1</i>	putative transcriptional activator 1
5	54329	57188	<i>nysR2</i>	putative transcriptional activator 2
	57180	59961	<i>nysR3</i>	putative transcriptional regulator 3
	60367	61045	<i>nysR4</i>	putative response regulator
10	61736	62495	<i>nysR5</i>	putative repressor
	63615	62553 C	ORF2	putative transcriptional regulator
	63765	64959	ORF1	putative peptidase (aminohydrolase)

15

nys 2 (SEQ ID No. 2)

	Start	End	Gene	Description
20				Name
	1191	456 C	<i>nysF</i>	putative 4'-phosphopantheteine transferase
	3092	1277 C	<i>nysG</i>	putative ABC transporter
25	4824	3072 C	<i>nysH</i>	putative ABC transporter
	5122	6154	<i>nysD3</i>	putative GDP-mannose-4,6- dehydratase
	6338	27541	<i>nysI</i>	nystatin PKS, modules 9-13 (incomplete)

30

"C" in the table above refers to complementary strands.

It will be appreciated that *nysD2*, *D1*, *A*, *B*, *C*, *E*, *R1* to *R5*, *ORF1*, *ORF2* of SEQ ID No. 1 and *nysF*, *G*, *H* and *D3*, and the partial *nysI* sequences of SEQ ID No. 2, correspond to their named counterparts in SEQ ID No. 35, which represents the whole complete coding sequence for the gene cluster, and comprises the nucleotide sequences

35

- 15 -

of SEQ ID Nos. 1 and 2. There is however a difference in the nucleotide numbering as between the corresponding features in SEQ ID Nos. 2 and 35; for SEQ ID Nos. 1 and 2, the first nucleotide in the stop codon is recognised as the end of the gene, whereas for SEQ ID No. 35, the third nucleotide of the stop codon is recognised as the end of the gene. *nysD1*, *D2* and *D3* are also known as *NysDI*, *DII* and *DIII*, and *nysR1* to *R5* are also known as *nysRI* to *RV*.

10 As regards SEQ ID No. 1 (*nys1*) and the gene sequence *nysRI*, Table 1 shows this to comprise nucleotides 51405 to 54303. Further sequence analysis has revealed however that there are in fact two start codons, nucleotides 51405-51407 encoding GTG and
15 nucleotides 51408-51410 encoding ATG. Accordingly, nucleotide 51408 of SEQ ID No. 1 represents an alternative start nucleotide for *nysRI*. ATG is preferred as a start codon to GTG, and consequently 51408 is regarded the start nucleotide in future
20 references to *nysRI*. The start of *nysRI* in SEQ ID No. 35 is indicated in Table 1 above as the ATG codon, and the translation product of *nysRI* shown in SEQ ID No. 9 below is deduced from nucleotides 51408-54303 of SEQ ID No. 1. [In an alternative presentation of the *NysRI*
25 translation product, wherein nucleotide 51405 of SEQ ID No. 1 is the start nucleotide, SEQ ID No. 9 is modified by the inclusion of an additional "first" amino acid, V].

 As regards SEQ ID Nos. 35 and 1, and the gene
30 sequence *nysRIV*, Table 1 shows this to comprise nucleotides 120676 to 121308 in SEQ ID No. 35 (*nysRIV* short). Further sequence analysis has revealed however that there are in fact two start codons, nucleotides 120676-120678 encoding GTG and nucleotides 120628-120630
35 encoding GTG. These start codons correspond to start codons in SEQ ID No. 1 for *nysR4* at nucleotides 60367-60369 (as stated in Table 1) and 60415-60417. The

- 16 -

upstream GTG (120628-120630 in SEQ ID No. 35, which corresponds to 60367-60369 in SEQ ID No. 1) is preferred as a start codon (see Example 6), and consequently 120628 is regarded the start nucleotide in future references to *nysRIV*. The start of *nysRIV* in SEQ ID No. 35 is indicated in Table 1 above (*nysRIV* short) as the downstream start codon 120676-120678, and the deduced translation product of *nysRIV* named herein as "NysRIV short" is shown in SEQ ID No. 12. Table 1 also shows the alternative start codon of *nysRIV* (*nysRIV* long) in SEQ ID No. 35 as the upstream start codon 120628-120630. Thus a preferred alternative presentation of the NysRIV translation product, named herein "NysRIV (long)", is shown in SEQ ID No. 43.

Alternative representative parts of the SEQ ID No. 35, or 1 and/or 2 sequences include the nucleotide sequences between the respective "start" and "end" nucleotide positions, either individually or collectively.

The translation products of the respective "genes" have been deduced and the amino acid sequences are set out in the following SEQ ID Nos. shown below:

	<u>Gene Product</u>	<u>SEQ ID No.</u>
25	NysD2 (NysDII) (partial)	3
	NysD1 (nysDI)	4
	NysA	5
	NysB	6
	NysC	7
30	NysE	8
	NysR1 (NysRI)	9
	NysR2 (NysRII)	10
	NysR3 (NysRIII)	11
	NysR4 (NysRIV) (short)	12
35	NysR4 (NysRIV) (long)	43
	NysR5 (NysRV)	13
	ORF2	14

- 17 -

	ORF1	15
	NysF	16
	NysG	17
	NysH	18
5	NysD3 (NysDIII)	19
	NysI	20
	NysDII (complete)	36
	NysI (complete)	37
	NysJ	38
10	NysK	39
	NysL	40
	NysM	41
	NysN	42

15 SEQ ID Nos. 4, 10, 12, 13, 14, 16, 18, 41 and 43 show valine (V) as the first amino acid. According to practice in this field and conceptual translation of bacterial DNA, the first amino acid of a protein (translation product) is always methionine (M),

20 regardless of the start codon. Accordingly, translation products and amino acid sequences of the present invention include not only SEQ ID Nos. 4, 10, 12, 13, 14, 16, 18, 41 and 43 as presented but also modifications of the aforesaid sequences in which the

25 first V is replaced with M. References to SEQ ID Nos. 4, 10, 12, 13, 14, 16, 18, 41 and 43 below will be understood to include not only the sequences as presented, but also the said sequences wherein the first V is replaced with M.

30 Viewed from an alternative aspect, the present invention also provides a nucleic acid molecule comprising a nucleotide sequence encoding one or more amino acid sequences selected from SEQ ID Nos 3 to 20 or 36 to 43, or a nucleotide sequence which is

35 complementary thereto or degenerate therewith.

Also provided are nucleic acid molecules comprising nucleotide sequences encoding one or more amino acid

- 18 -

sequences (i.e. polypeptides) which exhibit at least 60% sequence identity with any one of SEQ ID Nos. 3 to 20 or 36 to 43.

5 A further aspect of the invention provides a polypeptide encoded by a nucleic acid molecule of the invention as defined herein.

More particularly, this aspect of the invention provides a polypeptide comprising:

- 10 (a) all or part of an amino acid sequence as shown in any one or more of SEQ ID Nos. 3 to 20 or 36 to 43; or
(b) all or part of an amino acid sequence which has at least 60% sequence identity with any one or more of SEQ ID Nos. 3 to 20 or 36 to 43.

15 In particular the amino acid sequence may exhibit at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 98% identity with the polypeptide of any one of SEQ ID Nos. 3 to 20 or 36 to 43. Alternatively, the amino acid sequence may exhibit at least 70%, 75%, 80%, 85%, 90%, 95% or 98% similarity with the polypeptide of any one of
20 SEQ ID Nos. 3 to 20 or 36 to 43. Amino acid (polypeptide) sequences meeting the % sequence identity or similarity criteria herein are regarded as "substantially identical". The polypeptide of the invention may be an isolated, purified or synthesized
25 polypeptide. The term "polypeptide" is used herein to include any amino acid sequence of two or more amino acids i.e. both short peptides and longer lengths (i.e. polypeptides) are included.

Amino acid sequence identity or similarity may be
30 determined using the BestFit program of the Genetics Computer Group (GCG) Version 10 Software package from the University of Wisconsin. The program uses the local homology algorithm of Smith and Waterman with the default values: Gap creation penalty = 8, Gap extension
35 penalty = 2, Average match = 2.912, Average mismatch = - 2.003.

A "part" of the amino acid sequence of any one of

- 19 -

SEQ ID Nos. 3 to 20 or 36 to 43 (or of a "substantially identical" sequence as defined above) may comprise at least 20 contiguous amino acids, preferably at least 30, 40, 50, 70, 100, 150, 200, 300, 400, 500, 1,000, 2,000, 5,000 or 10,000 contiguous amino acids.

The polypeptide, and preferably also the part thereof, is functionally active according to the definitions given above, e.g. is enzymatically active or has a regulatory or transport functional activity. The part may not itself be functionally active but may in some instances provide regions with functional properties of the whole, e.g. represent the active site or co-factor binding site required for enzymatic activity.

The studies described in the Examples below have characterised the nucleotide and polypeptide sequences of the invention and various functional regions within them have been identified. Such functional regions form separate aspects of the invention. For the various translation products (i.e. gene products of SEQ ID Nos. 3 to 20 or 36 to 43), these functional regions are summarised below in Table 2 below:

Table 2: Molecule Features of Translation Products of
SEQ ID Nos. 3 to 20 and 36 to 43

(i) SEQ ID No.: 5
 Translation Product Name : NysA

	Start	End	Name	Description
	AA	AA		
	8	430	KS ^s	KS domain, loading module
35	528	841	AT	AT domain, loading module
	855	1055	DH	DH domain, loading module
	1285	1359	ACP	ACP domain, loading module

- 20 -

(ii) SEQ ID No.: 6

Translation Product Name: NysB

5	Start	End	Name	Description
	42	462	KS1	KS domain, module 1
	578	897	AT1	AT domain, module 1 (mMCoA-specific)
10	911	1110	DH1	DH domain, module 1 (inactive)
	1201	1447	KR1	KR domain, module 1
	1484	1559	ACP1	ACP domain, module 1
	1579	2004	KS2	KS domain, module 2
15	2117	2439	AT2	AT domain, module 2 (mMCoA-specific)
	2453	2659	DH2	DH domain, module 2 (inactive)
	2749	2996	KR2	KR domain, module 2
	3025	3102	ACP2	ACP domain, module 2
20				

(iii) SEQ ID No.: 7

Translation Product Name: NysC

25	Start	End	Name	Description
	35	455	KS3	KS domain, module 3
	546	858	AT3	AT domain, module 3
	872	1073	DH3	DH domain, module 3
30	1381	1628	KR3	KR domain, module 3
	1662	1735	ACP3	ACP domain, module 3
	1757	2180	KS4	KS domain, module 4
	2291	2603	AT4	AT domain, module 4
	2617	2818	DH4	DH domain, module 4
35	3124	3371	KR4	KR domain, module 4
	3407	3480	ACP4	ACP domain, module 4
	3501	3924	KS5	KS domain, module 5

- 21 -

	4032	4346	AT5	AT domain, module 5
	4360	4561	DH5	DH domain, module 5
	4953	5239	ER5	ER domain, module 5
	5248	5495	KR5	KR domain, module 5
5	5528	5601	ACP5	ACP domain, module 5
	5623	6046	KS6	KS domain, module 6
	6166	6478	AT6	AT domain, module 6
	6492	6704	DH6	DH domain, module 6
	7038	7281	KR6	KR domain, module 6
10	7315	7388	ACP6	ACP domain, module 6
	7408	7831	KS7	KS domain, module 7
	7939	8253	AT7	AT domain, module 7
	8267	8470	DH7	DH domain, module 7
	8812	9086	KR7	KR domain, module 7
15	9120	9193	ACP7	ACP domain, module 7
	9214	9637	KS8	KS domain, module 8
	9758	10072	AT8	AT domain, module 8
	10086	10289	DH8	DH domain, module 8
	10657	10904	KR8	KR domain, module 8
20	10939	11012	ACP8	ACP domain, module 8

(iv) SEQ ID No.: 9

Translation Product Name: Nys R1

25	Start	End	Name	Description
	42	49	P-loop	ATP/GTP binding site motif A
	904	932	HTH	LuxR-type helix-turn-helix motif (DNA binding)

30

(N.B. In the alternative representation of NysR1 above, where nt 51405 of SEQ ID No. 1 is regarded as the start codon, these start and amino acid end numbers would each increase by 1.)

35

- 22 -

(v) SEQ ID No.: 10

Translation Product Name: NysR2

5	Start	End	Name	Description
	902	930	HTH	LuxR-type helix-turn-helix motif (DNA binding)

10

(vi) SEQ ID No.: 11

Translation Product Name: NysR3

	Start	End	Name	Description
15				
	26	47	LZ	Leucine zipper motif (DNA binding)
	548	568	TM1	Transmembrane domain (putative)
	583	610	TM2	Transmembrane domain (putative)
20	884	912	HTH	LuxR helix-turn-helix motif

(vii) SEQ ID No.: 12

Translation Product Name: NysR4 (short)

25

	Start	End	Name	Description
	97	104	P-loop	ATP/GTP binding site motif A
30	149	177	HTH	LuxR helix-turn-helix motif (DNA binding)

- 23 -

(viii) SEQ ID No.: 13

Translation Product Name: NysR5

Start	End	Name	Description
6	40	HTH	DeoR helix-turn-helix motif (DNA binding)

10

(ix) SEQ ID No.: 14

Translation Product Name: ORF2

15

Start	End	Name	Description
186	202	HTH	AsnC HTH motif signature

(x) SEQ ID No.: 17

Translation Product Name: NysG

20

Start	End	Name	Description
31	313	TM	Transmembrane regions
392	399	P-loop	ATP/GTP binding site
25 496	510	ABC	ABC transporters signature

(xi) SEQ ID No.: 18

Translation Product Name: NysH

30

Start	End	Name	Description
17	288	TM	Transmembrane regions
368	375	P-loop	ATP/GTP binding motif A
35 472	486	ABC	ABC transporters signature

- 24 -

(xii) SEQ ID No.: 20

Translation Product Name: NysI (partial)

5	Start	End	Name	Description
	34	448	KS9	KS domain, module 9
	572	890	AT9	AT domain, module 9
	904	1123	DH9	DH domain, module 9
10	1443	1686	KR9	KR domain, module 9
	1720	1793	ACP9	ACP domain, module 9
	1813	2236	KS10	KS domain, module 10
	2346	2664	AT10	AT domain, module 10
	2678	2890	DH10	DH domain (inactive), module 10
15	2983	3229	KR10	KR domain, module 10
	3266	3339	ACP10	ACP domain, module 10
	3358	3780	KS11	KS domain, module 11
	3898	4217	AT11	AT domain, module 11
				(mMCoA-specific)
20	4231	4432	DH11	DH domain (inactive), module 11
	4523	4770	KR11	KR domain, module 11
	4806	4879	ACP11	ACP domain, module 11
	4801	5325	KS12	KS domain, module 12
	5432	5754	AT12	AT domain, module 12
25	5768	5977	DH12	DH domain (inactive), module 12
	6068	6315	KR12	KR domain, module 12
	6348	6421	ACP12	ACP domain, module 12
	6454	6873	KS13	KS domain, module 13

30 (xiii) SEQ ID No.: 37

Translation Product Name: NysI

	Start	End	Name	Description
35	34	448	KS9	KS domain, module 9
	572	890	AT9	AT domain, module 9
	904	1123	DH9	DH domain, module 9

- 25 -

	1443	1686	KR9	KR domain, module 9
	1720	1793	ACP9	ACP domain, module 9
	1813	2236	KS10	KS domain, module 10
	2346	2664	AT10	AT domain, module 10
5	2678	2890	DH10	inactive DH domain, module 10
	2983	3229	KR10	KR domain, module 10
	3266	3336	ACP10	ACP domain, module 10
	3355	3777	KS11	KS domain, module 11
	3898	4217	AT11	AT domain (methylmalony-CoA-specific), module 11
10	4231	4432	DH11	inactive DH domain, module 11
	4523	4769	KR11	KR domain, module 11
	4806	4879	ACP11	ACP domain, module 11
	4901	5325	KS12	KS domain, module 12
15	5432	5754	AT12	AT domain, module 12
	5768	5977	DH12	inactive DH domain, module 12
	6068	6315	KR12	KR domain, module 12
	6348	6421	ACP12	ACP domain, module 12
	6454	6873	KS13	KS domain, module 13
20	6973	7293	AT13	AT domain, module 13
	7307	7448	DH13	inactive DH domain, module 13
	7535	7774	KR13	inactive KR domain, module 13
	7813	7886	ACP13	ACP domain, module 13
	7908	8323	KS14	KS domain, module 14
25	8430	8741	AT14	AT domain, module 14
	8755	8962	DH14	inactive DH domain, module 14
	9050	9296	KR14	KR domain, module 14
	9319	9394	ACP14	ACP domain, module 14

30

(xiii) SEQ ID No.: 38

Translation Product Name: NysJ

35

Start	End	Name	Description
41	464	KS15	KS domain, module 15
578	889	AT15	AT domain, module 15

- 26 -

	903	1102	DH15	DH domain, module 15
	1446	1731	ER15	ER domain, module 15
	1740	1988	KR15	KR domain, module 15
	2023	2096	ACP15	ACP domain, module 15
5	2117	2538	KS16	KS domain, module 16
	2635	2953	AT16	AT domain, module 16
	2967	3167	DH16	inactive DH domain, module 16
	3257	3500	KR16	KR domain, module 16
	3539	3612	ACP16	ACP domain, module 16
10	3634	4057	KS17	KS domain, module 17
	4153	4472	AT17	AT domain, module 17
	4486	4725	DH17	inactive DH domain, module 17
	4997	5245	KR17	KR domain, module 17
	5277	5350	ACP17	ACP domain, module 17

15

(xiii) SEQ ID No.: 39

Translation Product Name: NysK

20

Start	End	Name	Description
34	457	KS18	KS domain, module 18
568	881	AT18	AT domain, module 18
898	1102	DH18	inactive DH domain, module 18
1416	1663	KR18	KR domain, module 18
1695	1769	ACP18	ACP domain, module 18
1849	2066	TE	thioesterase domain

30

(xiv) SEQ ID No.: 43

Translation Product Name: NysRIV (long)

35

Start	End	Name	Description
31	85	PAS	PAS-like domain
113	120	P-loop	ATP/GTP binding site motif A
165	193	HTH	LuxR helix-turn-helix motif (DNA binding)

- 27 -

As referred to in the above "inactive" denotes DH domains which lack the conserved amino acid sequence representing the active site motif $H(X_3)G(X_4)P$ found in DH in other PKSs. It will however be appreciated that
5 these domains may have activity although this is likely to be distinct from the activity of DH domains in other PKSs.

It will be seen that SEQ ID Nos. 5 (NysA), 6 (NysB), 7 (NysC), 20 and 37 (NysI), 38 (NysJ) and 39
10 (NysK) constitute actual "PKS" enzymes, namely enzymes involved in polyketide synthesis. These gene products contain identifiable enzymatic domains and modules which are tabulated in Table 2 above and shown also in Figures 4, 8 and 9 (see also Example 1 and Table 4, and Example
15 4 below which describes the DNA sequence analysis of nystatin biosynthesis gene cluster in more detail). Such individual domains and molecules, as identified herein form separate aspects of the present invention.

SEQ ID NOS 3 and 26 (NysDII), 4 (NysDI), 8 (NysE),
20 16 (NysF), 19 (NysDIII), 40 (NysL), 41 (NysM) and 42 (NysN) represent other enzymes functional in polyketide or macrolide synthesis e.g. in polyketide release from PKS, post-translational PKS modification, and polyketide modification. SEQ ID NOS 10 to 15 and 43 (NysRI to
25 NysRV, and ORF2) respectively represent transcriptional regulators, and SEQ ID NOS 17 and 18 (NysG and NysH) represent transport proteins which are presumed to be involved in polyketide transport from the cell. This is also described in more detail in the Examples below.
30 Such functional proteins represent separate aspects of the present invention. Also included are functional parts or fragments of such proteins i.e. active parts or fragments which retain (i.e. exhibit measurable levels of) the biological activity of the parent molecule from
35 which they are derived (i.e. of the whole protein or polypeptide).

The nucleotide sequences of the present invention

- 28 -

provide important tools and information which can be utilised in a number of ways to manipulate nystatin biosynthesis, to synthesise new nystatin derivatives or novel polyketide or macrolide structures, and to provide
5 novel or modified PKS systems (by "PKS system" here is meant a polyketide synthesis system i.e. a gene cluster or protein complex, collection or assembly, which is functional in polyketide synthesis, but which is not necessarily restricted to PKS enzymes or enzymatic
10 domains, and which may contain also other functional activities, e.g. other enzymatic (e.g. modificatory) or transporter or regulatory functional proteins).

Thus, for example, the entire nystatin PKS gene cluster or PKS synthetic system as provided herein, or a
15 portion thereof, may be subjected to modification so as to modify one or more genes, or one or more modules, or enzymatic domains, or functional sequences within it. Such modified or derivatised PKS systems may be used to synthesize novel or modified polyketide moieties, as
20 will be described in more detail below. In this situation, the nystatin PKS system provided herein, or a fragment or portion thereof, may function as an "origin" or "template" or "source" system or sequence for modification.

25 More particularly, in one such embodiment and as further described below, the non-functional parts (e.g. non-biologically active parts) of said system may be utilised as a "scaffold", and the functional parts (e.g. sequences encoding enzymatic portions) may be modified
30 to yield the derivative or modified PKS system. In some embodiments only a single selected, or few selected functional (e.g. enzymatic) regions may be modified, leaving the remaining sequence or structure largely intact.

35 Alternatively, the functional portions may be utilised as tools or materials for the modification of other "scaffold" structures e.g. individual nystatin

- 29 -

genes, modules or domains may be used for introduction (e.g. insertion or replacement) into other PKS scaffold structures e.g. PKS scaffold systems derived from PKS systems for other macrolide antibiotics e.g. erythromycin, rapamycin etc.

5 Included within the scope of the invention are synthetic or recombinant polyketide synthase enzymes derived from the scaffold encoded by the nystatin gene cluster which are modified to include one or more
10 functional units derived from other modular enzymes. Such functional units may encode a catalytic or transport protein domain for example a ketoreductase domain from a PKS enzyme or an ACP domain from a modular hybrid polyketide/peptide synthesising enzyme. Such
15 domains can be derived from enzyme domain DNA sequences from, for example, polyketide synthesising enzymes, peptide synthesising enzymes, hybrid peptide polyketide synthesising enzymes, fatty acid synthesising enzymes or other enzyme domains known in the art. Analogously,
20 there are included within the scope of the invention, synthetic or recombinant polyketide synthase enzymes derived from the scaffold encoded by a different polyketide synthase gene cluster, or modular enzyme encoding gene cluster, which are modified to include one
25 or more functional units derived from the nystatin gene cluster.

Thus, the sequence and activity information provided here for the nystatin biosynthesis gene cluster may be used to alter existing known gene clusters and
30 hence the products they produce. In particular, selection and incorporation of particular domains described herein (or modification of existing sequences) into existing PKS gene clusters will allow incorporation of particular properties attributable to the nystatin
35 gene cluster.

Thus, in a very general sense, the present invention provides the use of the nucleic acid molecules

- 30 -

of the invention as defined herein in the preparation of a modified PKS system, or in the preparation of modified polyketide molecules.

Such novel or modified polyketide or macrolide
5 molecules form a separate aspect of the present invention.

The nucleotide sequences may be utilised in this way according to the present invention in a random or directed or designed manner, e.g. to obtain and test a
10 particular predetermined or pre-designed structure, or to create random molecules, for example libraries of polyketide structures, e.g. for screening (this is also described in more detail below).

Whether for modification within the nystatin-PKS
15 scaffold, or for introduction into an alternative scaffold structure, the genes or genetic elements which can be modified include not only the actual PKS genes (which encode NysA, NysB, NysC, NysI, NysJ and NysK) or the individual molecules or domains thereof, but also
20 genes encoding other enzymes or functional proteins involved in nystatin biosynthesis and transport (referred to herein collectively as "PKS genes" or "nystatin genes").

As regards the actual PKS genes, as will be
25 described in more detail below, these may be modified to change the nature of the loading domain molecule which determines the nature of the starter unit, the number of modules, the nature of the extender, as well as the various dehydratase, reductase and synthase activities
30 which determine the structure of the polyketide chain.

Other genes which can be modified include the thioesterase gene, (encoding NysE; SEQ ID NO. 8), which may be modified to increase the efficiency of the PKS system (in the case of a thioesterase having "editing"
35 activity which clears the inappropriate substrates from the PKS). If the thioesterase simply cleaves the final product off the PKS, it can be used for making nystatin

- 31 -

derivatives with a smaller macrolactone ring by truncating nystatin PKS, and fusing this thioesterase to the end of the truncated protein via genetic engineering.

5 Regulatory genes: activators can be overexpressed and repressors inactivated in order to boost polyketide or antibiotic production. This may be of particular importance for the production of new nystatin derivatives in recombinant strains, (which may be
10 produced in very small quantities).

 The putative 4'-phosphopantetheine (PPT) transferase gene (encoding NysF; SEQ ID NO: 16) can be overexpressed in order to achieve efficient post-translational modification and full functionality of the
15 PKS. It can also be used for expression of the nystatin (or other) PKS in a heterologous host, which lacks the specific PPT activity. Such hosts may include *E.coli*, *Saccharomyces cerevisiae*, etc.

 Deoxysugar genes: glycosyltransferase (encoding
20 NysDI; SEQ ID NO:4) can be overexpressed in order to boost glycosylation of the synthesised molecules e.g. novel nystatin derivatives. It can also be modified by *in vitro* mutagenesis in order to increase its specificity towards the new substrates. Inactivation of this
25 glycosyltransferase will result in a recombinant strain producing non-glycosylated nystatin (probably also lacking some modifications) which can be used, for example, for chemical modifications, or enzymatic assays for screening new modification activities.

30 Aminotransferase (NysDII; SEQ ID Nos: 3 and 36) may be inactivated to give a nystatin derivative. This enzyme is presumed to attach the amino group on the deoxysugar mycosamine. This gene may also be expressed in other streptomycetes in order to achieve the same
35 reaction with another deoxysugar normally lacking an amino group.

 ABC transporters (e.g. NysH and NysG; SEQ ID NOs 17

- 32 -

and 18): can be overexpressed in order to make the efflux of nystatin and its derivatives more efficient. They may also be mutated in order to shift their specificity towards different compounds. They may be inactivated, if it is desired for any reason to accumulate the nystatin or its derivatives inside the cell.

The genes encoding monooxygenases NysL and NysN (SEQ ID Nos. 40 and 42) can be inactivated in *S. noursei* in order to obtain non-hydroxylated and non-oxidized nystatin derivatives. Alternatively, they can be mutated with the aim of changing their specificities toward nystatin precursors. Overexpression of nysL and nysN may potentially lead to increased yield of nystatin or its derivatives if the hydroxylation and/or oxidation steps are limiting in the nystatin biosynthetic pathway. Genetic manipulations with nysM encoding ferredoxin (SEQ ID No.41) might also be useful if one wishes to establish an *in vitro* P450 hydroxylase system for modifications of nystatin precursors.

Thus, in addition to modification of the nystatin PKS system, or modification of other PKS systems by using the "nystatin" genes, the nucleic acid molecules of the invention can also be utilised to manipulate or facilitate the biosynthetic process, for example by extending the host range or increasing yield or production efficiency etc.

In order to enable practice of the invention according to the principles above, the invention also provides an expression vector, and host cells containing a nucleic acid molecule as herein defined.

Also provided are methods for production of a polyketide or macrolide molecule (e.g. nystatin or a nystatin derivative), comprising expressing within a host cell, a nucleic acid molecule as defined above. The polyketide or macrolide molecule produced within the host cell or secreted or exported by the host-cell as a

- 33 -

result of expression of the nucleic acid molecule (i.e. expression of the introduced "PKS synthesis machinery") may then be recovered.

5 This method of the invention may thus involve growing or cultivating the host cell under conditions whereby the nucleic acid molecule is expressed, and allowing the expression product(s) of the nucleic acid molecule to synthesise the polyketide/macrolide molecule, or in other words, growing or cultivating the
10 host cell under conditions wherein the polyketide or macrolide is produced.

Also provided are methods for preparing recombinant nucleic acid molecules according to the invention, comprising inserting the nucleic acid molecules
15 containing the nucleotide sequences of the invention into another nucleic acid molecule, e.g. into vector nucleic acid, e.g. vector DNA.

Expression vectors of the invention may include appropriate control sequences such as for example
20 translational (e.g. start and stop codons, ribosomal binding sites) and transcriptional control elements (e.g. promoter-operator regions, termination stop sequences) linked in matching reading frame with the nucleic acid molecules of the invention.

25 Vectors according to the invention may include plasmids and viruses (including both bacteriophage and eukaryotic viruses) according to techniques well known and documented in the art, and may be expressed in a variety of different expression systems, also well known
30 and documented in the art.

A variety of techniques are known and may be used to introduce such vectors into prokaryotic or eukaryotic cells for expression, or into germ line or somatic cells to form transgenic animals. Suitable transformation or
35 transfection techniques are well described in the literature.

The invention also includes transformed or

- 34 -

transfected prokaryotic or eukaryotic host cells, for transgenic organisms containing a nucleic acid molecule according to the invention as defined above. Such host cells may for example include prokaryotic cells such as
5 *E.coli*, *Streptomyces* and other bacteria, eukaryotic cells such as yeasts or the baculovirus-insect cell system, transformed mammalian cells and transgenic animals and plants.

The nucleic acid molecules contained in the
10 expression vectors, and host cells and organisms etc. above may also be, as will be described in more detail below, derivative nucleic acid molecules, derived from the nucleic acid molecules defined above, either by modification or by introducing said molecules or parts
15 thereof into, or combining with, other nucleic acid molecules.

Thus, in one aspect, the invention provides recombinant materials for the production of combinatorial libraries of polyketides wherein the
20 polyketide members of the library are synthesized by modified PKS systems derived from the naturally occurring nystatin A1 system provided herein by using this system as a scaffold. Generally, many members of these libraries may themselves be novel compounds, and
25 the invention further includes novel polyketide members of these libraries. The invention methods may thus be directed to the preparation of an individual polyketide. The polyketide may or may not be novel, but the method of preparation permits a more convenient method of
30 preparing it. The resulting polyketides may be further modified to convert them to antibiotics, typically through glycosylation. The invention also includes methods to recover novel polyketides with desired binding activities by screening the libraries of the
35 invention.

Thus, in one aspect, the invention is directed to a method of preparing a nucleic acid molecule which

- 35 -

contains or comprises a nucleotide sequence encoding a modified polyketide synthase enzyme or enzyme system. The method comprises using the nystatin PKS encoding sequence as provided herein as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic or other functional activities e.g. by mutagenesis, inactivation (e.g. by deletion or insertion), or replacement. The thus modified nucleotide sequence encoding a modified PKS can then be used to modify a suitable host cell and the cell thus modified employed to produce a polyketide different from that produced by the native nystatin PKS, whose scaffolding has been used to support modifications of enzymatic activity.

Alternatively, one or more portions of the nucleotide sequence that encode enzymatic or other functional activities may be introduced into an alternative (i.e. different "second") PKS scaffold (i.e. a scaffold derived from a further "second" PKS system, different from the nystatin PKS system).

The invention is also directed to polyketides thus produced and the antibiotics to which they may then be converted.

In another aspect, the invention is directed to a multiplicity of cell colonies comprising a library of colonies wherein each colony of the library contains an expression vector for the production of a different modular PKS, but derived from the nystatin PKS of the invention, as defined above. The library of different modular PKS may be obtained by modifying one or more of the regions of a naturally occurring "nystatin" gene or gene cluster encoding an enzymatic activity so as to alter that activity, leaving intact the scaffold portions of the naturally occurring gene.

In another aspect, the invention is directed to a multiplicity of cell colonies comprising a library of colonies wherein each colony of the library contains a

- 36 -

different modular PKS derived from the nystatin PKS of the invention. The invention is also directed to methods to produce libraries of PKS complexes and to produce libraries of polyketides by culturing these colonies, as well as to the libraries so produced. In addition, the invention is directed to methods to screen the resulting polyketide libraries and to novel polyketides contained therein.

As mentioned above, a structural and functional sequence analysis of the nystatin PKS gene cluster/system is presented in the Examples below and the DNA sequences are shown in SEQ ID NO 1, 2 and 35, and further analysed in Tables 1 and 2 above, SEQ ID NOs 3 to 20, and 36 to 42 and in Figures 2, 3, 7, 8 and 9. The modular and "domain" encoding structure of the "PKS" gene may be seen. A module may typically contain a ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP). These three functions are sufficient to activate an extender unit and attach it to the remainder of the growing molecule. Additional activities that may be included in a module relate to reactions other than the Claisen condensation, and include a dehydratase activity (DH), an enoylreductase activity (ER) and a ketoreductase activity (KR). The loading module catalyses the initial condensation, i.e. it begins with a "loading domain" represented by AT and ACP, which determine the nature of the starter unit. The "finishing" of the molecule is believed to be regulated by thioesterase activity (TE) and it is believed that this is achieved by the TE activity embedded in NysK. This thioesterase appears to catalyse cyclization of the macrolide ring thereby increasing the yield of the polyketide product. The NysE TE activity is believed to be an "editing" one, and participates in cleaving off certain substrates from the nystatin PKS complex.

It will be seen from the sequences, Figures and

- 37 -

Tables above and below, that the regions in the genes and modules that encode enzymatic activities are separated by linker or "scaffold"-encoding regions. These scaffold regions encode amino acid sequences that space the enzymatic/functional activities at the appropriate distances and in the correct order. Thus, these linker regions collectively can be considered to encode a "scaffold" into which the various activities are placed in a particular order and spatial arrangement. It should however be noted that in some instances regions of the scaffold may be deleted or modified without significantly affecting the activity of the resultant PKS. Indeed the sequence encoding some domains with functional activity may be fully or partially deleted without significant effects. Thus as used herein "scaffold" refers to portions of PKS cluster not directly attributable to functional e.g. enzymatic activities of said PKS, but responsible for maintenance of its overall activity, e.g. providing correct spatial orientation or structure. Said scaffold may comprise all linker and non-functional regions of the PKS or a functionally active (i.e. retaining structural integrity) part thereof. This organization is similar in other naturally occurring modular PKS gene clusters.

The invention provides libraries or individual modified forms, ultimately of polyketides, by generating modifications in the nystatin PKS gene cluster so that the protein complexes produced by the cluster have altered activities in one or more respects, and thus produce polyketides other than the natural product of the PKS (i.e. nystatin A1). Novel polyketides may thus be prepared, or polyketides in general prepared more readily, using this method. By providing a large number of different genes or gene clusters derived from the naturally occurring nystatin PKS gene cluster, each of which has been modified in a different way from the native cluster, an effectively combinatorial library of

- 38 -

polyketides can be produced as a result of the multiple variations in these activities. Alternatively the nystatin PKS "functional regions" (e.g. genes, modules or domains) may be used, for introduction into a
5 "scaffold" obtained from another naturally occurring PKS system. The modified PKS encoding sequences and systems used in the present invention thus represent modular polyketide synthases "derived from" a naturally occurring nystatin PKS.

10 By a modular PKS "derived from" the nystatin PKS is meant a modular polyketide synthase (or its corresponding encoding gene(s)) that retains the scaffolding of all of the utilized portion of the naturally occurring gene. (Not all modules or genes
15 need be included in the constructs). On the constant scaffold, at least one enzymatic or functional activity is mutated, deleted or replaced, so as to alter the activity. Alteration results when these activities are deleted or are replaced by a different version of the
20 activity, or simply mutated in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or extender unit, and/or stereochemistry, and/or
25 chain length or cyclization and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a deleted activity is replaced, the origin of the replacement activity may come from a corresponding activity in a different
30 naturally occurring polyketide synthase or from a different region of the same PKS. Alternatively, such a "derived" modular PKS may incorporate one or more enzymatic or other functional activities (or their encoding nucleotide sequences) obtained or derived from
35 the nystatin PKS described herein, in the scaffolding of a second, different modular PKS (or its gene).

Modification or manipulation of the modular PKS may

- 39 -

involve truncation, e.g. gene or domain or module deletion or domain/gene/module swapping, addition or inactivation, which may involve insertion or deletion. Alternatively, random or directed modifications (i.e. mutations) may be made in the nucleotide sequence of the selected portion (e.g. in a gene/domain/module etc).

The derivative may contain preferably at least a thioesterase activity from the nystatin PKS gene cluster.

Advantageously, a polyketide synthase "derived from" the nystatin PKS may contain the scaffolding encoded by all or the portion employed of the nystatin synthase gene, contains at least two modules that are functional, preferably four or more modules and contains mutations, deletions, or replacements of one or more of the activities of these functional modules so that the nature of the resulting polyketide is altered. This definition applies both at the protein and genetic levels. Particular preferred embodiments include those wherein a KS, AT, KR, DH or ER has been inactivated or deleted or replaced by a version of the activity from a different PKS or from another location within the same PKS. Also preferred are derivatives where at least one noncondensation cycle enzymatic activity (KR, DH or ER) has been deleted or wherein any of these activities has been mutated so as to change the ultimate polyketide synthesized.

Thus, there are five degrees of freedom for constructing a polyketide synthase in terms of the polyketide that will be produced. First, the polyketide chain length will be determined by the number of modules in the PKS system. Second, the nature of the carbon skeleton of the PKS will be determined by the specificities of the acyl transferases which determine the nature of the extender units at each position -- e.g. malonyl, methyl malonyl, or ethyl malonyl, etc. Third, the loading domain specificity will also have an

- 40 -

effect on the resulting carbon skeleton of the polyketide. Thus, the loading domain may use a different starter unit, such as acetyl, propionyl, and the like. Fourth, the oxidation state at various positions of the polyketide will be determined by the dehydratase and reductase portions of the modules. This will determine the presence and location of ketone, alcohol, double bonds or single bonds in the polyketide.

Finally, the stereochemistry of the resulting polyketide is a function of three aspects of the synthase. The first aspect is related to the AT/KS specificity associated with substituted malonyls as extender units, which affects stereochemistry only when the reductive cycle is missing or when it contains only a ketoreductase since the dehydratase would abolish chirality. Second, the specificity of the ketoreductase will determine the chirality of any β -OH. Finally, the enoyl reductase specificity for substituted malonyls as extender units will influence the result when there is a complete KR/DH/ER available.

Thus, the modular nystatin PKS system permits a wide range of polyketides to be synthesized. As compared to the aromatic PKS systems, a wider range of starter units including aliphatic monomers (acetyl, propionyl, butyryl, isovaleryl, etc.), aromatics (aminohydroxybenzoyl), alicyclics (cyclohexanoyl), and heterocyclics (thiazolyl) are found in various macrocyclic polyketides. Recent studies have shown that modular PKSs have relaxed specificity for their starter units. Modular PKSs also exhibit considerable variety with regard to the choice of extender units in each condensation cycle. The degree of β -ketoreduction following a condensation reaction has also been shown to be altered by genetic manipulation (Donadio, S. et al. Proc. Natl. Acad. Sci. USA (1993) 90:7119-7123). Likewise, the size of the polyketide product can be varied by designing mutants with the appropriate number

- 41 -

of modules (Kao, C. M. et al. J. Am. Chem. Soc. (1994) 116: 11612-11613). Lastly, these enzymes are particularly well-known for generating an impressive range of asymmetric centres in their products in a highly controlled manner. The polyketides and antibiotics produced by the methods of the present invention are typically single stereoisomeric forms. Although the compounds of the invention can occur as mixtures of stereoisomers, it is more practical to generate individual stereoisomers using this system. Thus, the combinatorial potential within modular PKS pathways based on any naturally occurring modular, such as the nystatin, PKS scaffold is virtually unlimited.

In general, the polyketide products of the PKS must be further modified, typically by glycosylation, in order to exhibit antibiotic activity. Methods for glycosylating the polyketides are generally known in the art; the glycosylation may be effected intracellularly by providing the appropriate glycosylation enzymes or may be effected *in vitro* using chemical synthetic means.

The macrolide antibiotics, polyketide moieties of which are synthesised by modular PKSs, may contain any of a number of different deoxysugars. The nystatin molecule contains mycosamine deoxysugar moiety. Deoxysugar biosynthesis starts typically with glucose-1-phosphate and proceeds through the action of dTDP-glucose synthase and dTDP-glucose-4,6-dehydratase. The product of the latter, typically a dTDP-4,6-keto-6-deoxyglucose, is further subjected to at least two of the following reactions - epimerisation, isomerisation, reduction, dehydration, transamination, or methylation - to give a dTDP-D-deoxysugar. The latter is then attached to the macrolactone ring via the action of a glycosyltransferase, hence providing for the glycosylation of the macrolide compound.

Glycosylation can also be effected using the non-glycosylated macrolides as starting materials, and using

- 42 -

mutants of streptomycetes or other organisms (i.e. *Myxococcus*, *Pseudomonas*, *Mycobacterium* etc.) that can provide the glycosylation activities. Alternatively, glycosyltransferase-encoding genes from the organisms mentioned above can be introduced in *S. noursei* or other organisms containing the native or modified nystatin PKS genes or portions thereof in order to provide the desired glycosylation. The deoxysugar biosynthesis genes from the nystatin gene cluster can be used for complementation of corresponding activities in different PKS producers, as well as for engineering the biosynthetic pathways for alternative deoxysugars.

The derivatives of nystatin PKS can be prepared by manipulation of the relevant genes, or by introducing the nystatin genes or portions thereof into another PKS. A large number of modular PKS gene clusters have been mapped and/or sequenced, including erythromycin, soraphen A, rifamycin, avermectin and rapamycin, which have been completely mapped and sequenced, and FK506 and oleandomycin which have been partially sequenced, and candicidin, pimaricin and nemadectin which have been mapped and partially sequenced. Additional modular PKS gene clusters are expected to be available as time progresses. These genes can be manipulated using standard techniques to delete or inactivate activity encoding regions, insert regions of genes encoding corresponding activities from the same or different PKS systems, or otherwise mutate using standard procedures for obtaining genetic alterations. Of course, portions of, or all of, the desired derivative coding sequences can be synthesized using standard solid phase synthesis methods such as those described by Jaye et al., J. Biol. Chem. (1984) 259:6331, and which are available commercially from, for example, Applied Biosystems, Inc.

In order to obtain nucleotide sequences encoding a variety of derivatives of the naturally occurring PKS, and thus a variety of polyketides for construction of a

- 43 -

library, a desired number of constructs can be obtained by "mixing and matching" enzymatic activity-encoding portions, and mutations can be introduced into the native host (i.e. nystatin) PKS gene cluster or portions thereof.

5 Mutations can be made to the native sequences using conventional techniques. The substrates for mutation can be an entire cluster of genes or only one or two of them; the substrate for mutation may also be portions of one or more of these genes. Techniques for mutation are well known in the art and described in the literature, for example in W098/49315 and the references cited therein. Such techniques include preparing synthetic oligonucleotides including the mutation(s) and inserting the mutated sequence into the gene encoding a PKS subunit using restriction endonuclease digestion. (See, e.g. Kunkel, T.A., Proc. Natl. Acad. Sci. USA (1985) 82:448; Geisselsoder et al. BioTechniques (1987) 5:786.) Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) which hybridizes to the native nucleotide sequence, at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located (Zoller and Smith, Methods Enzymol. (1983) 100:468). Primer extension is effected using DNA polymerase, the product cloned and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Selection can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations. See, e.g. Dalbie-McFarland et al. Proc. Natl. Acad. Sci. USA (1982) 79:6409. PCR mutagenesis will also find use for effecting the desired mutations.

Random mutagenesis of selected portions of the

- 44 -

nucleotide sequences encoding enzymatic activities can be accomplished by several different techniques known in the art, e.g. by inserting an oligonucleotide linker randomly into a plasmid, by irradiation with X-rays or ultraviolet light, by incorporating incorrect nucleotides during *in vitro* DNA synthesis, by error-prone PCR mutagenesis, by preparing synthetic mutants or by damaging plasmid DNA *in vitro* with chemicals. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, nitrosoguanidine, hydroxylamine, agents which damage or remove bases thereby preventing normal base-pairing such as hydrazine or formic acid, analogues of nucleotide precursors such as 5-bromouracil, 2-aminopurine, or acridine intercalating agents such as proflavine, acriflavine, quinacrine, and the like. Generally, plasmid DNA or DNA fragments are treated with chemicals, transformed into *E. coli* and propagated as a pool or library of mutant plasmids.

In addition to providing mutated forms of regions encoding enzymatic or other functional activity, regions encoding the desired functions or activities may be recovered from different locations in the same nystatin PKS, for example, using PCR techniques with appropriate primers. By "corresponding" activity encoding regions is meant those regions encoding the same general type of activity -- e.g. a ketoreductase activity in one location of a gene cluster would "correspond" to a ketoreductase-encoding activity in another location in the gene cluster.

If replacement of a particular target region in a host polyketide synthase is to be made (be this host nystatin PKS, or a different PKS into which "nystatin" sequences are to be inserted), this replacement can be conducted *in vitro* using suitable restriction enzymes or can be effected *in vivo* using recombinant techniques involving homologous sequences framing the replacement

- 45 -

gene in a donor plasmid and a receptor region in a recipient plasmid, genome or chromosome. Such systems, advantageously involving plasmids of differing temperature sensitivities are described, for example, in
5 PCT application WO 96/40968.

WO 00/77181 describes methods of assembling several DNA units in sequence into large DNA constructs which are applicable to the recombinant polyketide synthases within the scope of the invention.

10 The vectors used to perform the various operations to replace the enzymatic activity in the host PKS genes or to support mutations in these regions of the host PKS genes may be chosen to contain control sequences operably linked to the resulting coding sequences in a
15 manner that expression of the coding sequences may be effected in a appropriate host. However, simple cloning vectors may be used as well.

If the cloning vectors employed to obtain PKS genes encoding derived PKS lack control sequences for
20 expression operably linked to the encoding nucleotide sequences, the nucleotide sequences are inserted into appropriate expression vectors. This need not be done individually, but a pool of isolated encoding nucleotide sequences can be inserted into host vectors, the
25 resulting vectors transformed or transfected into host cells and the resulting cells plated out into individual colonies.

Suitable control sequences include those which function in eucaryotic and prokaryotic host cells.
30 Preferred hosts include prokaryotic hosts and fungal systems such as yeast, but single cell cultures of, for example, mammalian cells could also be used. There is no particular advantage, however, in using such systems. Particularly preferred are yeast and prokaryotic hosts
35 which use control sequences compatible with *Streptomyces* spp. Suitable control sequences for single cell cultures of various types of organisms are well known in

- 46 -

the art. Systems for expression in yeast, including control sequences which effect secretion are widely available and are routinely used. Control elements include promoters, optionally containing operator sequences, and other elements depending on the nature of the host, such as ribosome binding sites. Particularly useful promoters for prokaryotic hosts include those from PKS gene clusters which result in the production of polyketides as secondary metabolites, including those from aromatic (Type II) PKS gene clusters. Examples are act promoters, tcm promoters, spiramycin promoters, and the like. However, other bacterial promoters, such as those derived from sugar metabolizing enzymes, such as galactose, lactose (tac) and maltose, are also useful. Additional examples include promoters derived from genes encoding biosynthetic enzymes such as tryptophan synthase (trp), the β -lactamase (bla), bacteriophage lambda FL, T5 and T7. In addition, synthetic promoters, such as the tac promoter (U.S. Patent No. 4,551,433), can be used.

Other regulatory sequences may also be desirable which allow for regulation of expression of the PKS replacement sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

Selectable markers can also be included in the recombinant expression vectors. A variety of markers are known which are useful in selecting for transformed cell lines and generally comprise a gene whose expression confers a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium. Such markers include, for example,

- 47 -

genes which confer antibiotic resistance or sensitivity to the plasmid. Alternatively, several polyketides are naturally coloured and this characteristic provides a built-in marker for screening cells successfully transformed by the present constructs.

The various PKS nucleotide sequences, or a mixture of such sequences, can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, e.g. a single promoter. The PKS subunits or mixture of components can include flanking restriction sites to allow for the easy deletion and insertion of other PKS subunits or mixture components so that hybrid PKSS can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such as site-directed mutagenesis and PCR.

As described above, particularly useful control sequences are those which themselves, or using suitable regulatory systems, activate expression during transition from growth to stationary phase in the vegetative mycelium. The system contained in plasmid RM5, i.e. the actI/actIII promoter pair and the actII-ORF4, an activator gene, is particularly preferred (McDaniel et al., Science, v. 262, p 1546-1550, 1993). Particularly preferred hosts are those which lack their own means for producing polyketides so that a cleaner result is obtained. Illustrative host cells of this type include the modified *S. coeticotor* CH999 culture described in PCT application WO 96/40968 and similar strains of *S. lividans*.

The expression vectors containing nucleotide sequences encoding a modified PKS system or a variety of PKS systems for the production of different polyketides may then be transformed into the appropriate host cells, e.g. to construct a library. In one straightforward approach, a mixture of such vectors is transformed into

- 48 -

the selected host cells and the resulting cells plated into individual colonies and selected for successful transformants. Each individual colony will then represent a colony with the ability to produce a particular PKS synthase and ultimately a particular polyketide. Typically, there will be duplications in some of the colonies; the subset of the transformed colonies that contains a different PKS in each member colony can be considered the library. Alternatively, the expression vectors can be used individually to transform hosts, which transformed hosts are then assembled into a library. A variety of strategies might be devised to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the nystatin host gene cluster so that each colony in the library produces a different PKS and ultimately a different polyketide. The number of different polyketides that are produced by the library is typically at least three (e.g. 2 mutations in the PKS genes which may appear separately or in combination), more typically at least ten, and preferably at least 20, more preferably at least 50, reflecting similar numbers of different altered PKS gene clusters and PKS gene products. The number of members in the library is arbitrarily chosen; however, the degrees of freedom outlined above with respect to the variation of starter, extender units, stereochemistry, oxidation state, and chain length is quite large.

Methods for introducing the recombinant vectors of the present invention into suitable hosts are known to those of skill in the art and typically include the use of CaCl_2 or other agents, such as divalent cations, lipofection, conjugation, protoplast transformation and electroporation.

A wide variety of hosts can be used, even though some hosts natively do not contain the appropriate post-translational mechanisms to activate the acyl carrier

- 49 -

proteins of the synthases. These hosts can be complemented with the appropriate recombinant enzymes, for example NysF, to effect these modifications.

5 The polyketide producing colonies can be identified and isolated using known techniques and the produced polyketides further characterized. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity.

10 The libraries can thus be considered at four levels: (1) a multiplicity of colonies each with a different PKS encoding sequence comprising a different PKS cluster but all derived from the nystatin PKS cluster; (2) colonies which contain the proteins that
15 are members of the PKS produced by the coding sequences; (3) the polyketides produced; and (4) antibiotics derived from the polyketides.

Colonies in the library can be induced to produce the relevant synthases and thus to produce the relevant
20 polyketides to obtain a library of candidate polyketides. The polyketides produced can be screened for antimicrobial, antitumour, antihelminthic or immunosuppressive activities, as well as for binding to desired targets, such as receptors, signalling proteins,
25 and the like. The supernatants or culture pellets per se can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening methods involve detecting the binding of each member of the library to receptor or
30 other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art.

Alternatively, individual polyketide members of the
35 library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included.

- 50 -

A large number of novel polyketides may be prepared according to the method of the invention, as illustrated by the representative in the Examples below. These novel polyketides may be useful intermediates in
5 formation of compounds with antibiotic activity through glycosylation or reactions as described above. As indicated above, the individual polyketides may be reacted with suitable sugar derivatives to obtain compounds of antibiotic activity. Antibiotic activity
10 can be verified using typical screening assays such as those set forth in Lehrer, R. et al., J. Immunol. Meth. (1991) 137:167-173.

Thus, in a further aspect, the invention provides a method of preparing a nucleic acid molecule comprising a
15 nucleotide sequence encoding a modified nystatin PKS, wherein said modified nystatin PKS is derived from a nystatin PKS as defined herein (i.e. a naturally occurring nystatin PKS) encoded by a nucleotide sequence as defined herein) containing first regions which encode
20 enzymatic or other functional activities and second regions which encode scaffolding amino acid sequences, said method comprising

- (a) modifying at least one said first region; or
- (b) incorporating at least one said first region
25 into a scaffolding-encoding second region from a different PKS-encoding nucleotide sequence.

As discussed above in relation to the nucleic acid molecules of the invention, the first region may be any part (e.g. encoding a domain or a module or a part
30 thereof) of a nucleotide sequence of the invention (i.e. SEQ ID NO 1 or 2 or 35).

Also provided are (i) a method of preparing a modified nystatin PKS as defined above, said method comprising expressing a nucleic acid molecule prepared
35 as defined above within a host cell (i.e. culturing or growing a host cell containing such a nucleic acid molecule) under conditions whereby the modified nystatin

- 51 -

PKS is expressed), and (ii) a modified nystatin PKS so produced or so obtainable.

New polyketides produced by such a modified PKS are also within the scope of the invention, as are new antibiotics which are generally the glycosylated forms of these polyketides although some have activity without glycosylation which may be due to different post-translation modifications such as hydroxylation or oxidation.

The invention will now be described in more detail in the following non-limiting Examples with reference to the drawings in which:

Figure 1 shows the structure of the polyene antifungal antibiotic nystatin A1;

Figure 2 presents physical and functional maps of the *E.coli* - *Streptomyces* shuttle vector pSOK101, pSOK201 and pSOK804 used in Examples 1, 2 and 3;

Figure 3 is a schematic representation showing two regions of the *S.noursei* ATCC11455 genome encoding the nystatin biosynthesis gene (corresponding to SEQ ID NOS 1 and 2). Overlapping recombinant phages containing the presented DNA sequences are shown over the regions drawings (see Example 1);

Figure 4 is a schematic representation showing the functional organisation of the nystatin PKS NysA (SEQ ID NO 5), NysB (SEQ ID NO 6), NysC (SEQ ID NO 7) and NysI (SEQ ID NO 20) and their roles in nystatin biosynthesis;

Figure 5 is a schematic representation showing genetic manipulations of the module 5 in NysC PKS leading to production of new polyene compounds by recombinant *S.noursei* strains (see Example 2);

Figure 6 shows the UV spectra (in DMSO) for nystatin and new polyene compounds S44 and S48 obtained from recombinant *S. noursei* strains with genetically altered NysC PKS (see Example 2);

Figure 7 is a schematic representation showing the region of the *S. noursei* ATCC11455 genome encoding the

- 52 -

nystatin biosynthetic gene cluster (corresponding to SEQ ID No. 35). Gene organisation within the gene cluster is shown. The inserts from the overlapping recombinant phages encompassing the cloned region are shown above the physical/genetic map. The *nys* genes are designated with capital letters in *italic*, other ORFs are numbered;

Figure 8 is a representation showing a proposed model for nystatin biosynthesis in *S. noursei*; and

Figure 9 is a schematic representation showing the functional organisation of the nystatin PKS NysA (SEQ ID No. 5), NysB (SEQ ID No. 6), NysC (SEQ ID No. 7), NysI (SEQ ID No. 37), NysJ (SEQ ID No. 38) and NysK (SEQ ID No. 39) proteins. KS^S - ketosynthase with the Cys to Ser substitution in active site; KS - ketosynthase; AT - acetate-specific acyltransferase; mAT - propionate-specific acetyltransferase; DH - dehydratase; DHi - inactive dehydratase; ER - enoyl reductase; KR - ketoreductase; KRi - inactive ketoreductase; ACP - acyl carrier protein.

Figure 10 shows compounds that can be theoretically produced from the following manipulations within the nystatin gene cluster

- insertion of ER domain into module 3 (1)
- insertion of ER domain into module 4 (2);
- 25 -simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 3 (3);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 4 (4);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 7 (5);
- 30 -simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 8 (6);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 9 (7);
- 35 -simultaneous inactivation of the ER domain in module 5 and insertion of the ER domains into modules 8 and 9 (8);

- 53 -

-simultaneous inactivation of the ER domain in module 5 and insertion of the ER domains into modules 7 and 8 (9).

5 Figure 11 shows compounds that can be theoretically produced from the following manipulations within the nystatin gene cluster:

- replacement of methylmalonyl-specific acetyltransferase (AT) domain in module 11 of the nystatin PKS with malonyl-specific AT domain (10);
- 10 -replacement of malonyl-specific AT domain in module 12 with methylmalonyl-specific AT domain with simultaneous replacement of methylmalonyl-specific AT domain in module 11 with malonyl-specific AT domain (11);
- replacement of malonyl-specific AT domain in module 10
- 15 with methylmalonyl-specific AT domain with simultaneous replacement of methylmalonyl-specific AT domain in module 11 with malonyl-specific AT domain (12);
- inactivation of P450 monooxygenase-encoding genes nysL or nysN (whichever is found to be responsible for
- 20 oxygenation of the methyl group at C-16 on the nystatin molecule) (13).

Figure 12 shows compounds that can be theoretically produced from the following manipulations within the nystatin gene cluster:

- 25 -inactivation of dehydratase (DH) domain in module 3 of the nystatin PKS (14);
- inactivation of DH domain in module 4 (15);
- inactivation of DH domain in module 3 with simultaneous inactivation of ER domain in module 5 (16);
- 30 -inactivation of DH domain in module 4 with simultaneous inactivation of ER domain in module 5 (17) ;
- inactivation of DH domain in module 7 with simultaneous inactivation of ER domain in module 5 (18);
- inactivation of DH domain in module 8 with simultaneous
- 35 inactivation of ER domain in module 5 (19);
- inactivation of DH domain in module 9 with simultaneous inactivation of ER domain in module 5 (20).

- 54 -

Example 1 - Cloning of the nystatin biosynthesis gene cluster

Bacterial strains, plasmids and growth conditions

Bacterial strains and plasmids used in this study are listed in Table 3. New strains and plasmids developed in the course of this study are described herein and shown in Figure 3. *S.noursei* ATCC 11455 and its mutants were grown on solid ISP2 medium (Difco), and in liquid TSB medium (Oxoid). Intergeneric conjugation from *E.coli* ET12567 (pUZ8002) into *Streptomyces* strains was done as described in Flett et al, FEMS Microbial Lett., v. 155: 223-229, (1997), except for the "heat shock" time, which was reduced to 5 minutes. Apramycin (Fluka) at a concentration 50 mg/ml was used to select for the *S.noursei* transconjugants on solid medium. For inoculation of the *S.noursei* ATCC11455 transconjugants prior to total DNA isolation, liquid medium TSB supplemented with 20 µg/ml apramycin was used. *E.coli* strains were grown and transformed as described in Sambrook et al, Cold Spring Harbor Laboratory (1989). *E.coli* ET12567 (pUZ8002) was maintained on the media containing 20µg/ml chloramphenicol and 50 µg/ml kanamycin.

Table 3 - Bacterial strains, plasmids and phages used in this study

	Strain, plasmid or phage	Properties	Source or reference
	<i>E.coli</i> DH5α	general cloning host	Sambrook et al, 1989, supra
30	<i>E.coli</i> XL1-Blue MRA (P2)	host for the gene library	Stratagene
	<i>E.coli</i> ET12567	strain for intergeneric conjugation	MacNeil et al, Gene, v. 111: 61-68, 1992
	<i>S.noursei</i> ATCC 11455	wild type, nystatin producer	ATCC

- 55 -

	pGEM3Zf(-)	ColE1 replicon, Ap ^R , 3.2kb	Promega
	pGEM72f(-)	ColE1 replicon, Ap ^R , 3.0 kb	Promega
5	pGEM11Zf(-)	ColE1 replicon, Ap ^R , 3.2kb	Promega
	pUZ8002	RK2 derivative, Km ^R , Tc ^R	D.H.Figurski
	pWHM3	ColE1+pIJ101 replicons, Thio ^R , 7.2kb	Vara et al, J.Bacteriol v. 171: 5872-5881, 1989
	pSET152	ColE1 replicon+ ϕ C31 int, oriT, Am ^R , 5.5kb	Bierman et al, Gene, v. 116: 43- 49, 1992
	pSOK101	pWHM3 derivative in which the 3.1kb BamHI/SphI fragment was replaced with the 3.0 kb BamHI/SphI fragment from pSET152 containing ColEI, oriT and Am ^R , 7.1kb	This work (see Figure 2)
10	pGM11	pSG5 replicon, Neo ^R , 5.3kb	Wohlleben & Muth, In "Plasmids, a practical approach", Ed. Hardy, IRL Press, p 147-175, 1993
	pSOK201	pGM11 derivative in which the 1.2 kb EcoRI/HindIII fragment was replaced with the 3.0 kb EcoRI/HindIII fragment from pSOK101 containing ColEI, oriT and Am ^R , 7.1 kb	This work (see Figure 2)

- 56 -

	DASHII	bacteriophage λ vector	Stratagene
5	pSOK804	ColEI replicon, Am^R , OriT, int_{WVB} , $AttP_{WVB}$	This work (see Figure 2) and Van Mellaert et al., Microbiol., v. 144:3351- 3358, 1998
10	pGEM7ermELi	pGEM7ZF plasmid containing $PermE^*$ promoter	C.R. Hutchinson

15

Am-apramycin, Ap-ampicilli, Neo-neomycin, Thio-
thiostrepton, Km-kanamycin, Tc-tetracycline

Analysis of the secondary metabolite production

20 Fermentations were performed in Applicon 3-1
fermenters containing initially 1.3 l SAO-23 medium
containing (gl^{-1}): glucose H_2O , 90; NH_4NO_3 , 2.5; corn flour,
3.0; $MgSO_4 \cdot 7H_2O$, 0.4; KH_2PO_4 , 0.2; $CaCO_3$, 7; trace element
solution, 3ml. Trace element solution (gl^{-1}): $FeSO_4 \cdot 7H_2O$,
25 5.0; $CuSO_4 \cdot 5H_2O$, 0.39; $ZnSO_4 \cdot 7H_2O$, 0.44; $MnSO_4 \cdot H_2O$, 0.15;
 $Na_2MoO_4 \cdot 2H_2O$, 0.01; $CoCl_2 \cdot 6H_2O$, 0.02; HCl, 50. The
fermentations were performed at $28^\circ C$ with pH controlled
at 6.5-7.0 by HCl (2M) and NaOH (2M). The dissolved
oxygen was controlled at >40% of saturation by the
30 agitation (300-900 rpm) and aeration (0.25 vvm).
Inocula for the fermentations (3 vol-%) with SAO-23
medium were grown in TSB-medium (TSB, Oxoid CM129, 37
 gl^{-1}) at $28^\circ C$ in shake flasks (500-ml baffled Erlenmeyer
flasks with 100 ml medium; 200 rpm). Each shake flask
35 was inoculated with 0.2 ml spore suspension and
incubated for 18-20 hours. Nystatin production was
assayed by HPLC of the dimethylformamide extracts of the
cultures after fermentations (Raatikainen, J.
Chromatogr., v. 588, 356-360, (1991)).

40

- 57 -

DNA Manipulation

Plasmid, phage and total DNA preparations, endonuclease digestions, ligations and fractionation were performed as described previously (Sambrook et al., 1989, supra; Hopwood et al, 1985, supra). DNA fragments were isolated from agarose gels using the QIAGEN Kit (QIAGEN GmbH), labelled with the use of the digoxigenin kit from Boehringer Mannheim and used for Southern blot analysis according to the manufacturer's manual. DNA sequencing was performed at QIAGEN GmbH (Germany), and the data were analysed with the GCG software (Devereux et al, Nucleic Acids Res. v. 12: 387-395, 1984).

PCR-assisted amplification and cloning of a PKS-encoding DNA fragment from the *S.noursei* ATCC11455 genome

In order to obtain the DNA encoding the nystatin biosynthesis genes, the *S.noursei* ATCC1455 gene library was probed with labelled PKS-encoding DNA. To obtain the DNA probe, degenerate oligonucleotide primers were designed, which correspond to conserved amino acid regions within β -ketoacyl synthase (KS) and acyl carrier protein (ACP) domains of known modular PKSs. The degenerate primers used for amplification corresponded to the conserved amino acid motifs in ACP and KS domains in known PKS, and were designed according to the codon usage table for *Streptomyces* (Wright & Bibb, Gene, v. 113: 55-65, 1992). The ACP nucleotide primer (sense) corresponded to the motif Glu(Asp)Leu Gly Phe(Leu, Val) Asp Ser Leu (SEQ ID NO:21) and had the sequence 5' - GAG/C CTG/C GGC/G T/CTG/C GAC TCC/G CTG/C-3 (SEQ ID NO: 22). The KS nucleotide primer (antisense) corresponded to the motif Val Asp Thr Ala Cys Ser Ser (SEQ ID NO: 23) and had the sequence 5' - G/CGA G/CGA G/ACA/ G/CGC C/GGT GTC G/CAC-3' (SEQ ID NO: 24). Total DNA isolated from the *S.noursei* ATCC11455 was used as a template for polymerase chain reaction (PCR)-assisted amplification of the DNA fragment from the genome of this organism

- 58 -

with the use of KS and ACP oligonucleotide primers. From the relative position of the motifs on the modular PKSs it was assumed that resulting PCR product would be approx 0.7 kb in size. The 50 μ l PCR mixture contained
5 0.1 μ g of *S.noursei* ATCC11455 total DNA, 25 pm of each ACP and KS oligonucleotide primers dNTPs (final concentration 350 μ m), 1xPCR buffer from Expand High Fidelity PCR System (Boehringer Mannheim), and 1.5 U of
10 the DNA polymerase mixture from the same system. The PCR was performed on the Perkin Elmer GeneAmp PCR System 2400 with the following program: 1 cycle of denaturation at 96°C (4 min), 35 cycles of denaturation/annealing/synthesis at 94°C (45 sec) and 70°C (5 min) and 1 cycle
15 of final annealing/extension at 72°C (7 min). The 0.7 kb DNA fragment obtained with this procedure was cloned in pUC18 vector in *E.coli* DH5 α with the use of Sure Clone Ligation Kit (Pharmacia). One of the resulting recombinant plasmids, pPKS72 of 3.5kb, was subjected to DNA sequence analysis.

20 Subsequent cloning in *Escherichia coli* vector pUC18 and DNA sequencing of the resulting 0.7 kb PCR product, followed by conceptual translation and database search, confirmed that it encodes part of PKS type I. This DNA fragment was used as a probe for screening a *S.noursei*
25 ATCC11455 gene library constructed in the phage vector DASHIII (see below) and one recombinant phage, designated lambda DASHII-N1, which hybridized to the probe, was isolated. As described further below, DASHII-N1 was used to generate further probes.

30

Construction and screening of the *S.noursei* ATCC1455 gene library

The *S.noursei* ATCC11455 gene library was constructed in phage lambda vector DASHII (Stratagene)
35 according to manufacturer's instructions. Total DNA from *S.noursei* ATCC11455 was isolated as described in Hopwood et al (1985) supra, partially digested with Sau3

- 59 -

AI restriction enzyme and fractionated on the sucrose gradient as described in Sambrook et al (1989), supra. The fractions of *S.noursei* ATCC1455 DNA containing fragments of 13-17 kb in size were used for ligation with the DASHII vector arms digested with *Bam*HI restriction enzyme. *E.coli* XL1-Blue MRA (P2) (Stratagene) was used as a host for a gene library construction and propagation.

DNA fragments to be used as probes for screening the gene library were purified from agarose gels using QIAGEN Kit (QIAGEN GmbH, Germany), and labelled by the use of the digoxigenin (DIG) kit from Boehringer Mannheim (Germany), according to the manufacturer's instructions. Probes used for the library screening and relevant recombinant phages discovered:

Probe	DASHII recombinant phages found using this probe
PKS72	N1, N14
E12.1: from N1	N41, N42, N44, N45, N48
E4.7.2: from N1	N58
L42E9.1: from N42	N64, N76
B1.0_58: from N58	N69

Description of the probes:

1. PKS72 probe. The 0.7kb DNA fragment isolated from the pPKS72 plasmid (see above) with restriction enzymes *Eco*RI and *Hind*III was used as a PKS72 probe.
2. E12.1 probe. The 2.6 kb *Bam*HI fragment from the insert of recombinant phage N1, representing its left flanking region, was subcloned into pGEM3Zf(-), resulting in plasmid pGEM(B2.6)-1. This plasmid was digested with *Eco*RI/*Ava*I, and the 0.55 kb fragment corresponding to the left end of the N1 DNA insert was purified and used as E12.1 probe.
3. E4.7.2 probe. The 4.7 kb *Eco*RI fragment from the

- 60 -

recombinant phage N1, representing its right flanking region was subcloned into pGEM3Zf(-), resulting in plasmid pGEM (E4.7)-1. This plasmid was digested with *EcoRI/HindIII*, and the 1.5kb DNA fragment corresponding to the right end of the N1 DNA insert was purified and
5 named E4.7.2.

4. L42E9.1 probe. The 9.0 kb *EcoRI* fragment representing the left flanking region of the DNA insert in the recombinant phage N42 was subcloned into
10 pGEM3Zf(-), resulting in plasmid pH42E9.1. This plasmid was digested with *EcoRI/BamHI*, and the 0.6 kb fragment corresponding to the left flank of the N42 DNA insert was purified and used as L42E9.1 probe.

5. L58B1.0 probe. The 3.0 kb *EcoRI* fragment representing the right flanking region of the recombinant phage N58 DNA insert was subcloned into
15 pGEM3Zf(-), resulting in plasmid pGEM(E3.0)-58. From the latter plasmid, the 1.0 kb *BamHI* fragment which is located on the right end of the N58 DNA insert was
20 purified and used as L58B1.0 probe.

Gene disruption experiment with the nystatin biosynthesis gene cluster

A 4.2 kb *BamHI* DNA fragment isolated from the
25 DASHII-N1 recombinant phage was first cloned into the pGEM3Zf(-) vector in *E.coli*, resulting in the plasmid pGEM 4.2-1. DNA sequences on both ends of the cloned fragment were determined and, after database search, were found to encode PKS type I. The *S.noursei* DNA
30 fragment cloned in pGEM4.2-1 was excised from this plasmid with restriction enzymes *EcoRI* and *HindIII*, and ligated with the 3.0 kb *EcoRI/HindIII* fragment from the vector pSOK201 (see Table 3 and Figure 3), and transformed into *E.coli*. Plasmid DNA designated
35 pKO(4.2)-1 was recovered from transformants, and then was transferred into *S.noursei* ATCC1455 by conjugation as described in Example 1 under "Bacterial strains,

- 61 -

plasmids and growth conditions".

Since pKO(4.2)-1 is not capable of replicating in *S.noursei* ATCC11455, it was assumed that it will function as a suicide vector integrating into the genome of *S.noursei* via homologous recombination. As a result of such recombination, a PKS gene in the *S.noursei* ATCC11455 genome, for which 4.2 kb *Bam*HI fragment cloned in pKO(4.2)-1 was presumed to be internal, would have been inactivated by disruption of its coding region.

Integration of pKO(4.2)-1 into the genome of the three *S.noursei* transconjugants was confirmed by Southern blot analysis with the use of labelled 4.2 kb *Bam*HI fragment from pGEM4.2-1 as a probe. One of the *S. noursei* disruption mutants carrying pKO(4.2)-1 integrated into its genome was tested for nystatin production in parallel with the parental strain ATCC11455 (see above for methods under "analysis of secondary metabolite production". While the latter was shown to produce nystatin at the expected level, no nystatin production was detected in the pKO(4.2)-1 disruption mutant, thus confirming the requirement of the identified PKS for nystatin biosynthesis.

Cloning of the nystatin biosynthesis gene cluster

In order to clone the entire gene cluster for the nystatin biosynthesis, the DNA fragments derived from the ends of the *S. noursei* DNA insert in the recombinant phage DASHII-N1 (N1), and subsequently found overlapping recombinant phages, were used as probes for screening the gene library (see above for probes). This screen resulted in isolation of the recombinant phages N14, N41, N42, N44, N45, N48, N58, N64, N69, and N76 comprising two regions (SEQ ID NOs 1 and 2 respectively) of the *S. noursei* ATCC11455 genome (approx. 98 kb total), as depicted in Fig. 4. A gene disruption experiment with the 4.3 kb *Eco*RI/*Bam*HI DNA fragment derived from the recombinant phage N64 (performed

- 62 -

essentially the same way as described above), confirmed that the second region (SEQ ID NO. 2) also encodes nystatin biosynthesis genes.

5 *DNA sequence analysis of the nystatin biosynthesis gene cluster*

The complete DNA inserts from recombinant phages mentioned above were subcloned either as *Xba*I or *Eco*RI fragments into pGEM3Zf(-) vector in an *E. coli* host, and
 10 nucleotide sequences were determined on both DNA strands of these fragments. Computer-assisted analysis of the DNA sequences comprising the two regions of the nystatin biosynthesis gene cluster (SEQ ID NOs 1 and 2 respectively) resulted in identification of the genes
 15 shown on Fig. 4 and listed in Table 4.

Table 4. Genes identified within the nystatin biosynthesis gene cluster of *S.noursei*

20	Designation	Product	Putative function
	<i>nysR1</i>	transcriptional activator	regulation of nystatin production
	<i>nysR2</i>	transcriptional activator	regulation of nystatin production
	<i>nysR3</i>	transcriptional activator	regulation of nystatin production
	<i>nysR4</i>	transcriptional activator	regulation
25	<i>nysR5</i>	transcriptional repressor	regulation
	ORF1	peptidase	peptide metabolism
	ORF2	transcriptional activator	regulation
	<i>nysA</i>	PKS type I	nystatin polyketide backbone synthesis (loading module)
	<i>nysB</i>	PKS type I	nystatin polyketide backbone synthesis (modules 1&2)

- 63 -

	<i>nysC</i>	PKS type I	nystatin polyketide backbone synthesis (modules 3-8)
	<i>nysI</i> (incomplete)	PKS type I	nystatin polyketide backbone synthesis (modules 9-13)
	<i>nysE</i>	thioesterase	release of final product from PKS
5	<i>nysD1</i>	glycosyltransferase	attachment of mycosamine to the polyketide backbone
	<i>nysD2</i>	aminotransferase	mycosamine biosynthesis
	<i>nysD3</i>	GDP-mannose-4,6-dehydratase	mycosamine biosynthesis
	<i>nysH</i>	ABC transporter	efflux of nystatin
	<i>nysG</i>	ABC transporter	efflux of nystatin
10	<i>nysF</i>	4'-phosphopantetheine transferase	post-translational modification of PKS

Three complete (*nysA*, *nysB* and *nysC*) (in SEQ ID NO:1), and one incomplete (*nysI*) genes (in SEQ ID NO:2) encoding the PKSs type I were identified. The amino acid (aa) sequences of the products encoded by these four genes were analysed by comparison to the aa sequences of known PKSs type I (see also Table 2 above for molecule features). Since all four proteins displayed high degree of homology towards rifamycin and rapamycin PKSs (Aparicio et al., Gene, v. 169: 9-16, 1996; Tang et al., Gene, v. 216: 255-265, 1998), presumptive functional analysis of nystatin PKSs was based on comparison to the formers. The NysA protein of 1366 aa encodes one module of PKS composed of KS^S, AT, DH, KR, and ACP domains. The lack of a conserved cysteine residue in KS^S domain suggests that this module cannot perform condensation reaction, and thus most probably represents a loading module providing the acetate starter unit for initiation of nystatin

- 64 -

polyketide backbone biosynthesis (Bisang et al., *Nature*, v. 401: 502-505, 1999). Analysis of the 3192 aa sequence of NysB revealed that it contains two modules with KS, AT, KR, ACP, and apparently inactive DH domains. The AT domains identified in both NysB modules display features characteristic for the methylmalonylCoA-specific AT domains (Haydock et al., *FEBS Lett.*, v. 374: 246-248, 1995). This feature of the NysB protein suggests that it comprises 1st and 2nd modules involved in the nystatin polyketide backbone biosynthetic pathway, as the only two proximal methylmalonyl CoAs incorporated in nystatin molecule are the first two extender units. NysC protein of 11096 aa, the largest, to our knowledge, bacterial polypeptide discovered so far, is composed of 6 modules apparently responsible for the condensation steps 3 to 8 in nystatin polyketide chain biosynthesis (incorporation of C21 - C32). Module 5 of the NysC protein contains an ER domain, which is accountable for the reduction of the double bond between C29 and C28 (see Fig. 1). Besides module 5, all other modules in NysC are similar in that they all contain KS, AT, DH, KR and ACP domains. It was noticed that KR domains in modules 4 and 5 are 100% identical on the aa sequence level, and 99.9% identical at the level of DNA sequences encoding these domains. Thus, KR domains in modules 4 and 5 most probably represent an example of a relatively recent duplication in the process of evolution. NysI C-terminally truncated protein of 7066 aa, for which aa sequence information at the C-terminus is still missing, is composed of at least 4 complete modules. All these modules contain KS, AT, DH, KR and ACP domains, but the DH domains in three modules are apparently inactive, suggesting that the known part of NysI PKS is responsible for the elongation steps 9 through 12 (incorporation of C13 - C20). This assumption is further supported by the fact that the AT domain in

- 65 -

module 11 (based on sequence similarity) is specific for methylmalonylCoA extender, while all other AT domains in NysA, NysC and NysI are malonyl CoA-specific. Thus, the AT domain in module 11 is presumably responsible for the occurrence of the methyl group at C16 on the nystatin polyketide backbone, which is later oxidized to give a C16-coupled carboxyl (see Fig. 1).

Downstream of the *nysC* gene, a coding sequence *nysE* for thioesterase, presumably responsible for the release of mature polyketide chain from the nystatin PKSs, was found. The *nysR1*, *nysR2* and *nysR3* genes, were found downstream of *nysE*. The products of these genes are homologous to the presumed transcriptional regulators. NysR1 (966 aa), NysR2 (953 aa), and NysR3 (927 aa) proteins were all found to contain putative helix-turn-helix (HTH) DNA binding motifs of LuxR type at their C-termini. Beside that, NysR1 contained a distinct ATP/GTP binding motif, and NysR3 contained a "leucine zipper" (putative DNA binding) motif at their N-termini. The gene encoding NysR1 was found to contain a rare TTA codon (for Leu11) close to the beginning of the gene, suggesting that NysR1 expression might be regulated in *S. noursei* at the level of translation by a *bldA*-like gene (White & Bibb, J. Bacteriol, v. 179: 627-633, 1997).

In order to confirm the involvement of *nysR1* gene in the nystatin biosynthesis a gene disruption experiment was performed with the use of a 1379 bp *ApaI* DNA fragment internal for the *nysR1* coding sequence. The 1379 bp *ApaI* DNA fragment internal for the *nysR1* coding sequence and representing nt 51531-52910 of SEQ ID NO 1 was cloned into the *ApaI* site of the pGEM11Zf(-) vector, giving the recombinant plasmid pNRD1. The 1430 bp *EcoRI/HindIII* DNA fragment isolated from pNRD1 was ligated with the 3.0 kb *EcoRI/HindIII* fragment of pSOK201 resulting in the pNRD2 vector. The latter was subsequently used for *nysR1* gene disruption in *S.*

- 66 -

noursei ATCC11455 via conjugation from *E. coli* ET12567 (pUZ8002).

Analysis of the nystatin production by the *nysR1* disruption mutant revealed that it is not capable of producing nystatin. It shall be noted, however, that the phenotype observed for the *nysR1* disruption mutant, might reflect the polar effect of mutation on the *nysR2* and *nysR3* genes, which can be co-transcribed with *nysR1*.

Downstream of *nysR3*, the genes *nysR4*, *nysR5*, ORF1, and ORF2 were found. The *nysR4* gene product of 210 aa (this has subsequently found to be 266 aa expression from an upstream start codon at nucleotide 120628 in SEQ ID No. 35) shows similarity to transcriptional activators of response regulator type, and contains centrally located ATP/GTP binding and C-terminally located LuxR-type HTH DNA binding motifs. *NysR5* protein of 253 aa displays similarity to the transcriptional repressors, and contains a putative DeoR-type HTH DNA binding motif at its N terminus. It seems likely that *nysR4* and *nysR5* gene products are involved in regulation of nystatin biosynthesis based on their location proximal to the nystatin biosynthesis genes. ORF2, located downstream of *nysR5*, and transcribed in the opposite direction, encodes a 354 aa peptide showing similarity to transcriptional activators and having centrally located putative HTH DNA binding motif of AsnC type. Whether this gene is involved in regulation of nystatin biosynthesis is not apparent, as ORF1, located immediately upstream of ORF2, encodes a putative peptidase. It seems likely that ORF2 is rather involved in regulation of ORF1, but to confirm this, experimentation on ORF2 inactivation is required. The fact that the gene encoding a peptidase, for which no role in nystatin biosynthesis could be assigned, was found on the right flank of the sequenced region, suggests that the right border of the nystatin biosynthesis gene cluster had been identified.

- 67 -

On the left flank of the sequenced region encompassing the genes described above, two genes located upstream of *nysA*, and transcribed in the direction opposite to *nysA* were identified (Fig. 4).

5 The *nysD1* gene product of 506 aa displays considerable homology to the UDP-glucuronosyltransferases from mammals. This enzyme belongs to the UDP-glycosyltransferase family, and takes part in the process of elimination of potentially toxic xenobiotics

10 by the way of their glycosylation. It seems likely that *NysD1* represents a glycosyltransferase responsible for the attachment of the deoxysugar moiety (mycosamine) to the nystatin polyketide backbone. The product of *nysD2* shows a high degree of homology to the perosamine

15 synthetases and transaminases responsible for the attachment of amino groups to the deoxysugars in the process of their biosynthesis. Thus, *NysD2* presumably represents an aminotransferase involved in mycosamine biosynthesis.

20 Beside *nysI* encoding PKS type I, four other genes were identified within the second sequenced region (SEQ ID NO. 2) of the nystatin biosynthesis gene cluster of *S. noursei* ATCC11455 (Fig. 4). The 344 aa protein encoded by the *nysD3* gene is highly homologous to the

25 GDP-mannose-4,6-dehydratases required for deoxysugar formation. It is thus likely that the *NysD3* protein is involved in biosynthesis of the nystatin mycosamine moiety in *S. noursei* ATCC11455. *nysH* and *nysG* gene products of 584 aa and 605 aa, respectively, display a

30 high degree of similarity to transporters of the ABC family. Both *NysH* and *NysG* polypeptides contain a distinct ABC transporter signature at their C-termini, centrally located ATP/GTP binding motifs, and N-terminally located transmembrane regions. These two

35 proteins most probably are responsible for the ATP-dependent active efflux of nystatin from the producing organism, thus eliminating the danger of

- 68 -

membrane clogging by the hydrophobic nystatin molecules. The product of the *nysF* gene, a 245 aa polypeptide, displays homology to the 4'-phosphopantetheine transferases. The latter enzyme is responsible for the post-translational modification of the ACP domains of the PKSs, and is required for its full functionality (Cox et al., *FEBS Lett.*, v. 405: 267-272, 1997; Kealey et al., *Proc. Natl. Acad. Sci. USA*, v. 95: 505-509, 1998). It seems likely, therefore, that the NysF protein functions in post-translational modification of nystatin PKSs, and is required for nystatin biosynthesis.

Example 2 - Genetic manipulation of the nystatin PKS genes leading to production of novel polyene antibiotics

Nystatin belongs to the group of the polyene macrolide compounds, which are characterized by having 3 to 8 conjugated double bonds in their macrolactone ring. Nystatin itself is a tetraene, having 4 conjugated double bonds between C20 and C27. There is also a set of 2 conjugated double bonds on the nystatin molecule, between C30 and C33, which is separated from the set of 4 conjugated double bonds by C28-C29 (see Fig. 1). From the computer-assisted analysis of the NysC PKS it became apparent that the ER domain in module 5 in this protein is responsible for reduction of the double bond between C28 and C29. Thus, by inactivating this particular domain, it is theoretically possible to obtain a compound with a double bond between C28 and C29, thus joining two sets of conjugated double bonds in the nystatin molecule, and creating a heptaene macrolide compound.

To inactivate the ER domain in module 5 of NysC, the method of "in-frame" deletion within the *nysC* gene was chosen. The construction of the vector pERD4.2 for gene replacement in the NysC-encoding genomic region of

- 69 -

S. noursei was as follows:

Inactivation of ER domain in module 5 of NysC

PCR-assisted amplification of the 394 bp DNA
5 fragment representing the coding sequence for the C-terminal part of the ER domain in module 5 of *NysC*, and the coding sequence for the N-terminal part of the KR domain in module 5 (nt 32174 - 32559), was performed. The oligonucleotide primer ERD1 (5'-
10 GTTGGTACCCCACTCCCGGTCCGCAC-3', sense) (SEQ ID NO. 25) was selected from the nucleotide sequence of *nysC* gene comprising nt 32174-32190 (SEQ ID NO. 1) with additional nucleotides on the 5' end in order to create a *KpnI* restriction enzyme cleavage site. The oligonucleotide
15 primer ERD2 (5'-CCAGCCGCATGCACCACC-3', antisense (SEQ ID NO. 26)) was selected from the *nysC* coding DNA sequence, and comprised the DNA segment between nt 32559-32542 (SEQ ID NO. 1) containing a *SphI* restriction enzyme cleavage site. The resulting PCR fragment was digested
20 with *KpnI* and *SphI*, and ligated together with the 1828 bp *BamHI/KpnI* DNA fragment (nt 29224-31052 of SEQ ID NO. 1), and the 1273 bp *SphI/EcoRI* DNA fragment (nt 32548-33821 of SEQ ID NO. 1) into the *EcoRI/BamHI* - digested pGEM3Zf(-) vector. The ligation mixture was transformed
25 into the *E.coli* DH5 α , and recombinant plasmid of 6.7 kb designated pERD4.1 was recovered from one of the transformants. The latter contained a hybrid DNA fragment representing nt 29224-33821 of the SEQ ID NO. 1 DNA sequence with internal deletion between nt 31052 and
30 nt 32174 of the *nysC* coding region. This deletion eliminated the coding regions (aa 4837 to 5208 of *nysC*) for the part of the DH-ER interdomain linker and C-terminal part of the ER domain containing a putative NADP(H) binding site.
35 To construct the vector for inactivation of *NysC* ER4 domain in *S.noursei*, the recombinant DNA fragment was excised with *EcoRI* and *HindIII* restriction enzymes

- 70 -

from pERD4.1 and ligated with the 3.0 kb *EcoRI/HindIII* DNA fragment from pSOK201, and the ligation mixture was transformed into the *E.coli* DH5 α . Plasmid pERD4.2 of 6.5 kb was isolated from one of the transformants and
5 used to perform the gene replacement procedure in *S.noursei* ATCC11455 (see below). The recombinant *S.noursei* strains selected after this procedure were designated ERD44 and ERD48.

The latter plasmid was introduced into *S. noursei* ATCC11455 by intergeneric conjugation, and one
10 transconjugant, *S. noursei* (pERD4.2), was chosen for further manipulations. After the correct mode of integration of pERD4.2 into the genome of the *S. noursei* (pERD4.2) was confirmed by Southern blot analysis,
15 selection for the second crossover event was carried out as described in Sekurova et al., FEMS Microbiol Lett, v.177: 297-304, 1999 (and see below).

Gene replacement procedure

20 This method is carried out as described by Sekurova et al., 1999, supra.

The plasmid constructed for gene replacement as described above was introduced into *S.noursei* by conjugation from the *E.coli* ET 12567 (pUZ8002). One of
25 the clones carrying the plasmid integrated into the chromosome via homologous recombination was subjected to three rounds of sporulation on antibiotic-free ISP2 agar medium, and the progeny after the third round was tested for the loss of antibiotic resistance marker. Southern
30 blot analysis of the total DNA isolated from several antibiotic-sensitive strains with an appropriate probe was used to confirm the desired mutation.

Of 8 colonies, which had lost the selection marker, and thus undergone a second crossover event, 4 were
35 shown by Southern blot analysis to have reverted to the wild-type genotype. Two strains were shown to contain a large deletion apparently eliminating a substantial

- 71 -

portion of the nystatin gene cluster, while two other mutants contained either the desired 1116 bp deletion (ERD44), or a deletion which was somewhat larger than expected (ERD48). Analysis of the polyene antibiotic production by the strains ERD44 and ERD48 revealed that they do not produce nystatin. Instead, ERD44 was shown to produce a polyene compound having UV spectrum peaks characteristic for heptaens (Figure 6), as expected. Surprisingly, the ERD48 mutant was shown to produce hexaenic polyenes (according to spectroscopic analysis), which would be consistent with "in-frame" deletion of the complete module 5 from the NysC protein. In order to investigate the event which occurred in the ERD48 mutant in more detail, the DNA fragment was PCR-amplified from the genome of *S. noursei* ERD48 mutant, which would encompass the putative product of such deletion (see below).

PCR-assisted amplification and cloning of the PKS-encoding DNA fragment from the S.noursei ERD48 genome

The PCR reaction aimed at amplification of part of the mutant *nysC* gene in *S.noursei* ERD48 was carried out with oligonucleotide primers KR48.1 (sense, 5'-CCG CGT CGG ATC CGC CGA C-3') (SEQ ID NO: 27) and KR48.2 (antisense, 5'-AGC CTT CGA ATT CGG CGC C-3') (SEQ ID NO: 28) which corresponded to the nt 24744-24760 and nt33818-33833, respectively, of the DNA sequence in SEQ ID NO. 1. The 50 μ l PCR mixture contained 0.1 μ g of *S.noursei* ERD48 total DNA, 25 pm of each KR48.1 and KR48.2 oligonucleotide primers, dNTPs (final concentration 350 μ m), 1xPCR buffer from Expand High Fidelity PCR System (Boehringer Mannheim) and 1.5 U of the DNA polymerase mixture from the same system. The PCR was performed on the Perkin Elmer GeneAmp PCR System 2400 with the following program: 1 cycle of denaturation at 96°C (4 min), 35 cycles of denaturation/annealing/synthesis at 94°C (45 sec) and 70°C (10 min) and 1 full

- 72 -

cycle of final annealing/extension at 72°C (10 min).
 The 2.7 kb DNA fragment obtained with this procedure was
 digested with *EcoRI* and *BamHI* restriction enzymes, and
 ligated with pGEM3Zf(-) vector DNA digested with the
 same enzymes. The ligation mixture was introduced into
E.coli DH5 α by transformation, the plasmid pKR48 of
 5.9kb was isolated from one of the transformants and
 subjected to DNA sequence analysis.

The DNA sequence of the insert in pKR48 is present
 in SEQ ID NO. 29, (identified herein as ERD48 seq). The
 translation product is shown in SEQ ID NO. 30 and is a
 899 aa protein - the molecule features of SEQ ID NOs 29
 and 30 respectively are shown below:

15 SEQ ID NO: 29 (DNA; ERD48.seq)

	Start	End	Name	Description
	1	254	DH4	DH4 domain coding region, C-terminal
	1170	1913	KR4/5	hybrid ketoreductase domain, module 4/5
	2010	2231	ACP5	ACP5 domain coding region
20	2295	2700	KS5	KS5 domain coding region

SEQ ID NO: 30 (AA; translation product)

	Start	End	Name	Description
25	1	84	DH4	DH4 domain module 4, C- terminus
	390	637	KR4/5	hybrid KR domain
	670	743	ACP5	ACP domain, module 5
	764	899	KS5	KS domain, module 5, N- terminus

30 The cloning and DNA sequencing of this 2700 bp
 fragment confirmed that it encodes a part of a hybrid
 PKS module, which would be consistent with recombination

- 73 -

between DNA sequences encoding highly homologous KR domains in modules 4 and 5 of NysC. This recombination apparently has led to deletion of DNA sequences encoding C-terminal parts of KR and ACP domains of module 4, and
5 KS, AT, DH, ER domains, and N-terminal part of the KR domain of module 5, thus resulting in the loss of one complete module from NysC PKS.

Preliminary analysis of the compounds produced by the *S. noursei* ERD44 and ERD48 was carried out. It was
10 shown that a heptaenic compound produced by the ERD44 mutant has high antifungal activity against *Candida albicans*, an organism used in tests for antifungal activity. At this point it is not possible to accurately assay the activity of this compound
15 (tentatively named S44), because it is not yet properly purified, and its exact concentration is difficult to estimate. However, some rough estimates based on the UV absorbance at the wave lengths characteristic for nystatin, S44, and amphotericin (a heptaenic macrolide),
20 suggest that S44 might have 4-5 times higher antifungal activity compared to nystatin. HPLC analysis of the compounds produced by the ERD48 mutant suggests that at least 5 hexaenic macrolides with different retention times are produced by this strain (mixture called S48).
25 This probably reflects the different states of modifications of the macrolactone ring by i.e. glycosylation at C19, hydroxylation at C10, or oxidation of the methyl group at C16. This could have been expected, since reduction of the macrolactone ring size
30 most probably leads to the lower affinity of the modifying enzymes towards the new substrate.

Antifungal activity of the S48 mixture was tested, and found to constitute approx. 10% of nystatin activity. It seems probable that only one of the
35 compounds in the S48 mixture produced by ERD48, which is fully decorated by the ring-modifying enzymes, is responsible for the antifungal activity detected. Thus,

- 74 -

relative antifungal activity of this compound is impossible to assess prior to its purification. Some of the hexane antibiotics are known to have antibacterial, as well as antifungal activity (Ciftci, et al., J. Antibiot, v. 37: 876-884, 1984). It is thus possible that hexaenic compounds produced by the ERD48 mutant can be used for production of antibacterial agents. The changes in the NysC protein leading to production of new polyene compounds in ERD44 and ERD48 mutants, along with the predicted structures of their macrolactone rings are presented in Fig. 5.

Inactivation of DH8 domain in module 8 of NysC

The possibility of genetically manipulating the nystatin PKS was further exemplified by inactivation of the DH domain in module 8 of NysC. The plasmid pNPR1.1 for gene replacement within *nysC* gene, which would result in in-frame deletion of the DNA region encoding DH8 domain was constructed as below:

The 3989 bp *KpnI/BclI* DNA fragment (nt 43004-46993 of the region 1 DNA sequence (SEQ ID NO. 1) and the 2409 bp *BamHI/EcoRI* DNA fragment (nt 47680-50089 of the same) were excised from the DNA of recombinant phage N1 and ligated with vector pGEM3Zf(-) DNA digested with *EcoRI* and *KpnI*. The ligation mixture was transformed into *E.coli* DH5a, and recombinant plasmid pGEM-NPR1 was isolated from one of the transformants. The latter contained the hybrid DNA fragment representing the nt 43004-50089 of the region 1 DNA sequence (SEQ ID NO:1) with the internal deletion between nt 46993 and nt 47680. This deletion eliminated the DNA region encoding the aa 10150 to 10378 of the NysC polypeptide, thus affecting the DH8 domain in module 8, and DH8-KR interdomain linker in module 8. To construct the vector for inactivation of the NysC DH8 domain in *S.noursei*, the recombinant DNA fragment was excised with *EcoRI* and *HindIII* restriction enzymes from pGEM-NPR1 ligated with

- 75 -

the 3.0 kb *EcoRI*/*HindIII* DNA fragment from pSOK201 (see Fig. 3 and Table 3) and the ligation mixture was transformed into the *E.coli* DH5 α . Plasmid pNPR1.1 of 9.4 kb was isolated from one of the transformants and
5 used to perform the gene replacement procedure in *S.noursei* ATCC11455 according to Sekurova et al., 1999, supra.

The recombinant *S.noursei* strain selected after this procedure was designated NPR1.1 The *S. noursei*
10 NPR1.1 recombinant strain was shown by Southern blot analysis to contain the desired deletion in the DH8-coding sequence of *nysC*. Analysis of the secondary metabolites in the culture extracts of the *S. noursei* NPR1 recombinant strain by thin layer chromatography
15 (TLC) revealed the presence of presumed macrolide compounds. The relative mobility of these compounds differed from nystatin, and no UV spectra characteristic for nystatin could be detected in the extracts. It was suggested, that in the new molecule(s) produced by the
20 NPR1 recombinant a set of 4 double bonds on the nystatin aglycone has been disturbed, and that the macrolactone ring now contains a hydroxy group attached at C23 (Table 5). No attempts to purify the compound(s) from NPR1 were made, as the bioassay against *Candida albicans* made
25 with the NPR1 culture extracts showed very low antifungal activity. However, the NPR1 mutant can be potentially useful for further manipulations with the nystatin PKS.

30 **Inactivation of KR domain in module 7 (*NysC*)**

The 4404 bp DNA fragment was excised with *EcoRI* and *SmaI* restriction enzymes from the DNA of recombinant phage N1. The *EcoRI* site is situated in the polylinker of phage N1 to the left of the *S.noursei* DNA insert
35 starting at nt 38398 of the region 1 DNA sequence (SEQ ID NO. 1), while the *SmaI* site corresponds to nt 42802 of the region 1 DNA sequence. The 3303 bp DNA fragment

- 76 -

was excised with *Sma*I and *Bam*HI restriction enzymes from the DNA of the same recombinant phage. The *Sma*I and *Bam*HI restriction sites are situated at nt 43099 and nt 46402 respectively, of the region 1 DNA sequence (SEQ ID NO. 1). Both DNA fragments were ligated with the pGEM3Zf(-) DNA digested with restriction enzymes *Eco*RI and *Bam*HI, and the ligation mixture was transformed in *Escherichia coli* DH5 α . The plasmid pGEM-NPR2 of 10.7 kb was recovered from one of the transformants, which contained a recombinant DNA fragment which represented nt 38398-46402 of the region 1 DNA sequence (SEQ ID NO. 1) with the internal deletion between nt 42802 and nt 43099. This deletion results in elimination of the DNA region encoding aa 8753 to 8851 of the NysC protein encompassing a putative NADP(H) binding site in the KR domain of module 7. To construct the vector for inactivation of NysC KR7 domain in *S.noursei*, the recombinant DNA fragment was excised from pGEM-NPR2 with *Eco*RI and *Hind*III restriction enzymes, and ligated with the 3.0 kb *Eco*RI/*Hind*III DNA fragment from pSOK201, and the ligation mixture was transformed into the *E.coli* DH5 α . Plasmid pNPR2 of 10.7 kb was isolated from one of the transformants and used to perform the gene replacement procedure in *S.noursei* ATCC11455 (according to Sekurova et al., 1999, supra). The recombinant *S.noursei* strain selected after this procedure was designated NPR2.1.

Analysis of the secondary metabolites in the culture extracts of the *S. noursei* NPR2.1 recombinant strain by TLC revealed the presence of presumed macrolide compounds. The relative mobility of these compounds differed from nystatin, and from the metabolites produced by NPR1 mutant. No UV spectra characteristic for nystatin could be detected in the extracts. It is suggested, that in the new molecule(s) produced by the NPR2.1 recombinant a set of 4 double bonds on the nystatin aglycone has been disturbed, and

- 77 -

the macrolactone ring now contains a keto group at C25 (Table 5). No attempts at purification of the compound(s) from NPR2.1 were made, and no bioassays for antifungal activity with the NPR2.1 culture extracts were performed. This mutant has utility not only by virtue of the metabolites it produces, but also for further manipulation with nystatin PKS.

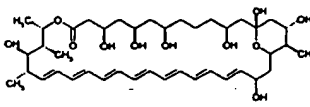
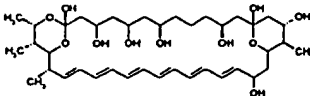
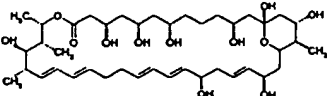
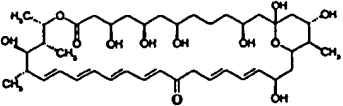
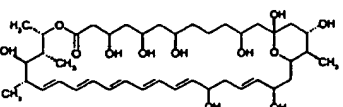
Inactivation of the ER domain in module 5 of the mutated NysC protein NysC_DH8)

To introduce the second mutation into the NysC protein with inactivated DH domain in module 8 (NysC_DH7), the plasmid pERD4.2 was introduced into the *S.noursei* mutant NPR1.1 and the gene replacement procedure was carried out as described in Sekurova et al., 1999, supra. This yielded recombinant *S.noursei* strain ERDH9 with mutations in both ER5 and DH8 coding sequences of *nysC*. The combination of these two mutations presumably leads to biosynthesis of the pentaenic nystatin derivative with a hydroxy group at C23 (Table 5). Preliminary analysis of the ERDH9 culture extracts confirmed that a polyene compound(s) is being produced by this strain although in quantities making identification of its true UV spectrum difficult. The preliminary data also show that this compound is preferentially accumulated in the culture supernatant, while nystatin produced by the wild-type *S. noursei* remains mostly associated with mycelium. This was consistent with the hypothesis that an additional hydroxy group on the nystatin molecule is responsible for increased water solubility of the compound(s). No attempts at purification of this new compound(s) were made, and no bioassays were performed.

To date, the Inventors have been unable to confirm the structure of the products of mutants NPR1.1, NPR2.1 and ERDH9.

- 78 -

Table 5. Predicted structures of nystatin derivatives produced via genetic engineering of *nysC* gene in *S. noursei* (see Example 2 for details)

Mutant	Expected structure (polyketide moiety only)	UV spectrum, nm (DMSO)	Activity
ATCC11455	nystatin (Fig. 1)	299, 312, 327	normal
ERD44		375, 395, 419, 444	high
ERD48		336, 352, 370, 391	low?
NPR1		-	low
NPR2.1		-	?
ERDH9		?	?

- 79 -

Example 3 - Manipulation of the regulatory genes leading to increased production of nystatin

Expression of nysR1 under the control of PerME promoter in S. noursei ATCC11455*

To further confirm that *nysR1* gene encodes a transcriptional activator of the nystatin biosynthesis genes, the latter was expressed in *S. noursei* ATCC11455 under the control of the PerME* promoter (Bibb et al., Mol. Microbiol., v. 14: 433-545, 1994). First, the plasmid pSOK804 for stable and efficient integration into the *S. noursei* ATCC11455 genome was constructed (Fig. 2). This was made by ligating together the 3.0 kb *SphI/HindIII* DNA fragment from pSET152 (Bierman et al., 1992, supra) and 2.3 kb *SphI/HindIII* fragment from bacteriophage VWB carrying functions necessary for site-specific integration (Van Mellaert et al., 1998, supra). Conjugation of pSOK804 from *E. coli* ET12567 (pUZ8002) into *S. noursei* ATCC11455 demonstrated that this plasmid integrates specifically into one site of the *S. noursei* genome at a frequency of $3 \cdot 10^{-6}$.

To clone the *nysR1* gene under PerME* promoter in pSOK804, the following procedure was employed. The N-terminal part of *nysR1* was PCR-amplified from the recombinant phage N1 DNA template with the oligonucleotide primers NR1.1 (sense) 5'-CGCCGCATGCTGTTCTACCCACGT-3' (SEQ ID NO: 31), and NR1.2 (antisense): 5'-GGCGGACCCGGTTCGGCCT-3' (SEQ ID NO: 32). The oligonucleotide primer NR1.1 sequence corresponded to the nt 51376-51391 of SEQ ID NO: 1 with addition of nucleotides CGCCGCATGC at the 5' end to create a site for *SphI* endonuclease. The oligonucleotide primer NR1.2 corresponded to the sequence complementary to nt 51964-51982 of SEQ ID NO: 1, and encompassed a restriction site for *AgeI* endonuclease. A 0.6 kb fragment was PCR-amplified using the phage N1 DNA as a template under conditions described in Example 1, and digested with *SphI*

- 80 -

and AgeI. The digested 0.6 kb fragment was ligated together with a 2.8 kb AgeI/EcoRI DNA fragment from the phage N1 insert (EcoRI site originating from the phage's polylinker) into pGEM7Zf(-) vector digested with SphI and EcoRI. The 2.8 kb fragment used in this ligation corresponded to nt 51971-54814 of SEQ ID NO: 1, and encompassed the C-terminal part of *nysR1* and N-terminal part of *nysR2* (encoding the first 162 aa of *NysR2*). The recombinant plasmid obtained as a result of this ligation was designated pNRE1. The 3.5 kb SphI/HindIII DNA fragment from pNRE1 was ligated together with the 0.3 kb EcoRI/SphI fragment from pGEM7ermELi (see Table 3) containing *PermE** promoter into pSOK804 vector digested with EcoRI/HindIII. The resulting plasmid, pNRE2, was introduced into the *S. noursei* ATCC11455 by conjugation (see Example 1) yielding recombinant strain *S. noursei* (pNRE2). Analysis of the nystatin production by the latter strain in shake-flasks with reduced glucose medium revealed that it produces 50% more nystatin compared to the wild-type strain, most likely due to overexpression of the *nysR1* gene from the *PermE** promoter. It therefore appears that *nysR1* gene encodes a positive activator that may be used for enhancing the production of nystatin and its derivatives in *S. noursei* strains.

Partial deletion of nysR5 gene in the S. noursei ATCC11455

To confirm the function of the *nysR5* gene predicted through analysis of its coding DNA sequence (see Example 1) in the regulation of nystatin biosynthesis, a specific mutation in *S. noursei* ATCC11455 genome was introduced. A DNA fragment from the *S. noursei* ACTT11455 genome encompassing nt 62037-63360 of the nucleotide sequence reported here was amplified by PCR with the primers NR5D1 (5'-GCGAGCGGCCGCTTCACCCCGCAACTCA-3') (SEQ ID NO: 33) and NR5D2 (5'-

- 81 -

CGCGAAGCTTGGCCGACTGCTCGACGTC-3') (SEQ ID NO: 34). The conditions for PCR were as described above. The 1341 bp PCR product was digested with *NotI* and *HindIII*, and ligated with the 1688 bp *EcoRI/NotI* DNA fragment (nt 5 60301-61989) and 3.0 kb *EcoRI/HindIII* fragment from pSOK201. The resulting plasmid, pNR5D, contained the *S. noursei* DNA fragment with a 43 bp deletion in the coding region of *nysR5* gene. This deletion creates a frame-shift mutation within the *nysR5* coding region, 10 subsequently leading to truncation of its product. As a result of such truncation, 165 C-terminal amino acids of NysR5 are eliminated, and replaced with 14 amino acids encoded by another reading frame (and thus unrelated to NysR5). The pNR5D plasmid was used to perform a gene 15 replacement procedure in *S. noursei* ATCC11455 as described in Sekurova et al., 1999, supra.

The mutation introduced through gene replacement led to a 5-15% increase in nystatin production by the resulting recombinant strain NR5D, compared to the wild-type *S. noursei*. Subtle but reproducible positive 20 effect of NysR5 C-terminal deletion on nystatin biosynthesis correlates well with the putative repressor function assigned to this protein based on computer analysis (see Example 1). Since the deletion introduced 25 in the *nysR5* gene does not eliminate the N-terminally located putative helix-turn-helix motif identified in this protein, the residual repressor activity of the truncated NysR5 polypeptide could account for the relatively small effect of this mutation on nystatin 30 production. Nevertheless, this result confirms the usefulness of introducing mutations in the repressor-encoding gene as further means for enhancing the production of nystatin and its derivatives.

- 82 -

Example 4 - Completion of the sequencing of the nystatin biosynthesis gene cluster

5 The DNA sequence spanning the gap between SEQ ID No. 1 and SEQ ID No. 2 was determined for both DNA strands on the overlapping inserts in recombinant phages N20, N32, N95, N98, and N99 covering this region.

10 Procedures used for sequencing, probe generation and screening were as described in Example 1. The probes used for library screening, and the relevant recombinant phages discovered were as follows:

	PKS 72	N20, N32
	L44ES3.5	N90
15	L76SN0.5	N95
	L20S0.64	N98, N99

Description of the probes:

20 6. L44ES3.5 probe. The 3.5 kb DNA fragment isolated from the phage N44 with restriction enzymes *EcoRI* and *SalI*, was used as a L44ES3.5 probe.

7. L76SN0.5 probe. The 0.5 kb DNA fragment isolated from the recombinant phage N76 with restriction enzymes *SacI* and *NotI*, was used as a L76SN0.5 probe.

25 8. L20S0.64 probe. The 0.64 kb DNA fragment isolated from the plasmid pL20EB3.7 with the restriction enzyme *SalI*, was used as a L20S0.64 probe.

30 New sequence information resulted in identification of complete *nysI* and *nysDII* genes, as well as the new genes *nysJ*, *nysK*, *nysL*, *nysM*, and *nysN* (see Figure 7). According to the new information, the NysI protein of 9477 aa (SEQ ID No. 37) represents a PKS composed of six modules, responsible for the elongation steps 9 to 14 of the nystatin polyketide backbone biosynthesis. All
35 these modules contain KS, AT, DH, KR and ACP domains. The presence of a *mAT* domain in module 11 is consistent with incorporation of methylmalonyl-CoA extender at this

- 83 -

elongation step. The DH domains in modules 10, 11, 12, and 14 seem to be inactive due the absence of the active site motif H(X₃)G(X₄)P. The KR domain in module 13 of NysI lacks the conserved motif aSRrG, and thus appears to be inactive. The latter feature, together with an inactive DH domain in module 11, most probably account for the presence of a six-membered ketalic ring (between C13 and C17) on the nystatin molecule.

The *nysJ* gene encoding a PKS, is located downstream of *nysI*, and is transcribed in the same direction. As judged from the organisation of modules in NysJ (SEQ ID No. 38), the latter is required for the elongation steps 15 to 17 in the nystatin macrolactone ring assembly. The DH domains in modules 16 and 17 within NysJ seem to be inactive, and the ER domain localized in module 15 is most probably responsible for the reduction of a double bond between C8 and C9.

The last, 18th module in the nystatin PKS system is represented by the NysK protein (SEQ ID No. 39) encoded by the *nysK* found downstream of *nysJ*. The NysK protein of 2066 aa is composed of KS, AT, inactive DM, ACP and TE domains. The NysK protein lacks a KR domain, and contains apparently inactive DH domain. A TE domain was identified at the C-terminus of NysK, suggesting that in addition to the condensation of the last extender unit, this protein also participates in the release of the mature nystatin polyketide chain from the PKS complex.

The gene organisation of the cluster is shown in Figure 9 and Figure 8 sets out the proposed involvement of the various proteins encoded in the nystatin biosynthetic pathway.

To confirm the involvement of *nysI*, and *nysJ* in nystatin biosynthesis, these genes were disrupted in *S. noursei* via homologous recombination using the conjugative suicide vectors pKNI1, and pKNJ1. The construction of the above vectors for disruption experiments was as follows. A 3.8 kb *EcoRI*/*BamHI*-

- 84 -

fragment internal for the *nysI*, and corresponding to nucleotides 23711-27541 of SEQ ID No. 35, was excised from N64 DNA and ligated into the corresponding sites of plasmid pGEM3Zf(-), resulting in plasmid pL64EB3.8. The
5 *S. noursei* fragment cloned into pL64EB3.8 was excised from this plasmid with restriction enzymes *EcoRI* and *HindIII*, and ligated with the 3.0 kb *EcoRI/HindIII*-fragment from the vector pSOK201, resulting in plasmid pKNI1.

10 A 3.7 kb *EcoRI/BamHI*-fragment internal for *nysJ*, and corresponding to nucleotides 43287-46992 of SEQ ID No. 35 was excised from phage N20 DNA, and ligated into the corresponding sites of plasmid pGEM3Zf(-), resulting in plasmid pL20EB3.7. The *S. noursei* fragment cloned
15 into pL20EB3.7 was excised from this plasmid with restriction enzymes *EcoRI* and *HindIII*, and ligated with the 3.0 kb *EcoRI/HindIII*-fragment from the vector pSOK201, resulting in plasmid pKNJ1. Both pKNI1 and pKNJ1 constructs were transformed into *E. coli* ET 12567
20 (pUZ8002), and further transferred into *S. noursei* ATCC 11455 by conjugation, as described in Zotchev et al., (2000) Microbiology 146: 611-619. No nystatin production was detected in either of the pKNI1 and pKNJ1 disruption mutants, thus confirming the requirement of
25 the identified PKS's for nystatin biosynthesis.

Three genes encoding proteins presumably involved in modification of the nystatin molecule were identified between *nysK* and *nysDII*. Both the *nysL* and *nysN* genes encode P450 monooxygenases of 394 aa and 398 aa,
30 respectively (SEQ ID Nos. 40 and 42 respectively), that are probably responsible for hydroxylation of the nystatin polyketide moiety at C10, and oxidation of the methyl group at C16. Which protein is responsible for which reaction is not yet clear, and additional
35 experiments are required for exact placement of *NysL* and *NysN* in the nystatin biosynthetic pathway. The *nysM* gene apparently encodes a ferredoxin of 64 aa (SEQ ID

- 85 -

No. 41), which presumably constitutes a part of one or both P450 monooxygenase systems, and serves as an electron donor [O'Keefe, D.P. & Harder, P.A. (1991). Occurrence and biological function of cytochrome P450 monooxygenases in the actinomycetes. *Mol. Microbiol.* 5, 2099-2105].

The DNA sequences extending the region depicted on Figure 7 (SEQ ID No. 35) approximately 10 kb to the left (recombinant phage N90), and approximately 5 kb to the right (recombinant phage N69) were determined. No genes with plausible functions in the nystatin biosynthesis were found, suggesting that the entire nystatin gene cluster had been identified.

Thus, based on the complete sequence information for the nystatin biosynthetic gene cluster, the following genes are identified and their roles described as follows (see also Figures 8 and 9):

Table 6. Genes identified within the nystatin biosynthetic gene cluster of *S. noursei* ATCC 11455

	Designation	Product	Putative function
25	<i>nysF</i>	putative 4'-phosphopanthetheine transferase	post-translational PKS modification
	<i>nysG</i>	ABC transporter	efflux of nystatin
	<i>nysH</i>	ABC transporter	efflux of nystatin
	<i>nysD3</i>	GDP-mannose-4,6-dehydratase	mycosamine biosynthesis
30	<i>nysI</i>	PKS type I	nystatin PKS (modules 9-14)
	<i>nysJ</i>	PKS type I	nystatin PKS (modules 15-17)
35	<i>nysK</i>	PKS type I	nystatin PKS (module 18 + TE)
	<i>nysL</i>	P450 monooxygenase	hydroxylation at

- 86 -

			C-10
	<i>nysM</i>	ferredoxin	electron transfer in P450 system
	<i>nysN</i>	P450 monooxygenase	oxidation of methyl group at C-16
5	<i>nysD2</i>	aminortransferase	mycosamine biosynthesis
	<i>nysD1</i>	glycosyltransferase	attachment of mycosamine
10	<i>nysA</i>	PKS type I	nystatin PKS (loading module)
	<i>nysB</i>	PKS type I	nystatin PKS (modules 1 and 2)
	<i>nysC</i>	PKS type I	nystatin PKS (modules 3-8)
15	<i>nysE</i>	thioesterase	release of polyketide chain from PKS
	<i>nysR1</i>	transcriptional activator	regulation of nystatin production
20	<i>nysR2</i>	transcriptional activator	regulation of nystatin production
	<i>nysR3</i>	transcriptional activator	regulation of nystatin production
25	<i>nysR4</i>	transcriptional activator	regulation
	<i>nysR5</i>	transcriptional repressor	regulation
	ORF2	transcriptional activator	regulation
30	ORF1	peptidase	peptide metabolism

- 87 -

Example 5 - Manipulation of nystatin biosynthetic genes:

Predictive examples presented below provide guidelines for the rational genetic manipulations of the nystatin biosynthetic genes aimed at specific chemical changes in the nystatin molecule. These manipulations are based on the current understanding of structure-function relationship of the polyene antibiotics the number of conjugated double bonds and the presence of two ionisable groups (exocyclic carboxyl and aminogroup belonging to the deoxysugar moiety mycosamine).

Changing the number and positions of conjugated double bonds.

The conjugated double bonds within the nystatin macrolactone ring are formed as a result of two reductive steps performed by a PKS modules with ketoreductase (KR), and dehydratase (DH) activities. Further reduction of the double bond can be brought about by introducing a enoylreductase (ER) activity in such PKS modules. This shall result in the formation of a completely saturated bond instead of a double bond at a specific step of nystatin biosynthesis. The following manipulations can be proposed (compounds that theoretically can be produced as a result of these manipulations are presented on Figure 10):

- insertion of ER domain into module 3 (1)
- insertion of ER domain into module 4 (2);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 3 (3);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 4 (4);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module

- 88 -

- 7 (5);
- simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 8 (6);
 - 5 - simultaneous inactivation of the ER domain in module 5 and insertion of the ER domain into module 9 (7);
 - simultaneous inactivation of the ER domain in module 5 and insertion of the ER domains into modules 8 and 9 (8);
 - 10 - simultaneous inactivation of the ER domain in module 5 and insertion of the ER domains into modules 7 and 8 (9);

15 These manipulations can be performed using the techniques for gene replacement described for *S. noursei* (Sekurova et al., FEMS Microbiol Lett, 177: 297-304, 1999). The materials for manipulations can be provided by the nystatin gene cluster itself, or other PKS gene clusters. The latter could be preferential from the point of view of genetic stability of the recombinant strains.

20

Removal/repositioning of the COOH group

25 Exocyclic carboxyl function of the polyene antibiotics is believed to play a crucial role in selective toxicity of these compounds. More specifically, the inter- and intramolecular interactions between the ionizable exocyclic carboxyl and amino group of micosamine moiety seem to be of particular importance (Resat, H., Sungur, F.A., Baginski, M., Borowski, E., Aviyente, V. 2000. Conformational properties of amphotericin B amide derivatives - impact on selective toxicity. Journal of Computer-Aided Molecular Design, v. 14, p 689-703). It is theoretically possible to either remove the exocyclic carboxyl from, or reposition it on

30

35

- 89 -

the nystatin molecule via manipulation of the nystatin biosynthetic genes. Nystatin derivatives produced via such manipulations could be useful on their own, or serve as substrates for further chemical modifications.

5 The following manipulations are proposed (resulting compounds are represented on Figure 11):

- replacement of methylmalonyl-specific acetyltransferase (AT) domain in module 11 of the nystatin PKS with malonyl-specific AT domain (10);
- 10 - replacement of malonyl-specific AT domain in module 12 with methylmalonyl-specific AT domain with simultaneous replacement of methylmalonyl-specific AT domain in module 11 with malonyl-specific AT domain (11);
- 15 - replacement of malonyl-specific AT domain in module 10 with methylmalonyl-specific AT domain with simultaneous replacement of methylmalonyl-specific AT domain in module 11 with malonyl-specific AT domain (12);
- 20 - inactivation of P450 monooxygenase-encoding genes *nysL* or *nysN* (whichever is found to be responsible for oxygenation of the methyl group at C-16 on the nystatin molecule) (13).

25 It shall be noted that specificity of the P450 monooxygenase responsible for the appearance of the exocyclic carboxyl function can be engineered so that it fulfills its function on the new substrates. Such methods as site-specific or random mutagenesis along with
30 error-prone PCR and DNA shuffling might prove useful for this purpose.

1.3. Introduction of additional hydroxyl functions (increasing water solubility).

35

Polyene antibiotics are very poorly soluble in water mostly due to a highly hydrophobic set(s) of

- 90 -

conjugated double bonds. Increasing water solubility can be an advantage in certain cases, as it expands pharmacologicxal properties of the drug (Golenser, J., Frankenburg, S., Ehrenfreund, T. & Domb, A.J. 1999.

5 Efficacious treatment of experimantal leishmaniasis with amphotericin B-arabinogalactan water-soluble derivatives. *Antimicrob Agents Chemother.*, v. 43: 2209-2214). To increase the water solubility of nystatin, we suggest to introduce additional hydroxyl functions (hydrophilic) to
10 the nystatin molecule. The following modifications of the nystatin biosynthetic genes can lead to the desired effect (resulting compounds depicted on Figure 12):

- inactivation of dehydratase (DH) domain in module 3 of the nystatin PKS (14);
- 15 - inactivation of DH domain in module 4 (15);
- inactivation of DH domain in module 3 with simultaneous inactivation of ER domain in module 5 (16);
- inactivation of DH domain in module 4 with
20 simultaneous inactivation of ER domain in module 5 (17) ;
- inactivation of DH domain in module 7 with simultaneous inactivation of ER domain in module 5 (18);
- 25 - inactivation of DH domain in module 8 with simultaneous inactivation of ER domain in module 5 (19);
- inactivation of DH domain in module 9 with simultaneous inactivation of ER domain in module 5
30 (20);

Extension, truncation or rebuilding of the macrolactone ring.

35 Novel derivatives of a polyene antibiotic nystatin can be obtained also through truncation of the nystatin PKS, leading to derivatives with a smaller macrolactone

- 91 -

ring (as exemplified in example 2 for the ERD48 mutant). This can be achieved through deletion of one or more modules from the nystatin PKS. Such truncations can lead to production of polyketides with 36- to 6-membered lactone rings that potentially can be useful for further modifications and synthesis of novel pharmaceuticals. Extension of the nystatin macrolactone ring can be achieved through insertion of additional modules into the nystatin PKS. Such manipulations can also lead to production of the lead compounds for pharmaceutical applications.

The nystatin molecule can be completely rebuilt by the way of shuffling the PKS modules between the nystatin, or other PKSs so that completely new derivatives are produced. In this respect, the method disclosed in the patent WO 00/77181 can prove useful for making the recombinant DNA constructs serving this purpose.

Finally, the nystatin biosynthetic genes can prove useful for manipulation of other macrolide antibiotic biosynthetic pathways. Both PKS and modification enzymes can prove useful for such purposes. It seems likely that nystatin biosynthetic genes will be most useful for manipulation of other polyene antibiotic biosynthetic clusters, such as the one for pimaricin (Aparicio, J.F., Fouces, R., Mendes, M.V., Olivera, N., Martin, J.F. 2000. A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. Chem Biol, v. 7: 895-905). High degree of similarity on the protein level between the nystatin and pimaricin biosynthetic enzymes will most probably ensure that their hybrids are functional. On the other hand, different specificities of the heterologous modification enzymes might provide new tools for further structural changes on the molecules produced by genetically engineered strains.

- 92 -

Example 6: Further manipulations of the regulatory genes leading to increased production of nystatin

Expression of *nysRII* under control of the *PermE** promoter in *S. noursei* ATCC 11455

To confirm the function of the *nysRII* gene predicted through analysis of its coding sequence (see Example 1), it was expressed in *S. noursei* ATCC 11455 under control of the *PermE** promoter. The 2168 bp *SalI*-*BclI* fragment from the phage N58 (representing C-terminal part of *nysRII*) was cloned into the *SalI*-*BamHI* digested pGEM11Zf(-), resulting in construct pC1A1. The 811 bp N-terminal part of the *nysRII* gene was PCR-amplified from the phage N58 template with the oligonucleotide primers NSR2.1 (5'-GCCGGCATGCGACGAA CAG GACGAGAGGT-3') (SEQ ID NO. 44.) and NSR2.3 (5'-GCCGTGGTCGACGAA GGC-3') (SEQ ID NO. 45). The conditions for PCR were as described above. The PCR fragment was digested with *SphI* and *SalI*, and ligated, together with the 2168 bp *SalI*-*HindIII* fragment from pC1A1 into *SphI*-*HindIII* digested pGEM3Zf(-) vector, giving the plasmid pC2A1. From the latter, the 3.0 kb *SphI*-*HindIII* fragment was isolated and ligated, together with the 0.3 kb *EcoRI*-*SphI* fragment containing the *PermE** promoter either with *EcoRI*-*HindIII* digested pSOK804 vector (generating plasmid pC3A1), or with the 3.0 kb *EcoRI*-*HindIII* fragment from pSOK201 (generating plasmid pC3E1). Since the pSOK804-based vectors integrate site-specifically in the *S. noursei* chromosome, the pC3A1 plasmid could be regarded as a construct for *nysRII* expression in-trans. Plasmid pC3E1, on the other hand, is a suicide vector capable of integrating into the *S. noursei* genome only via homologous recombination through the cloned *nysRII* gene, thus providing expression of the latter in-cis. Introduction of plasmids pC3A1 and pC3E1 resulted in recombinant strains C3A1 and C3E1, respectively. In the former strain *PermE** promoter is placed upstream of both *nysRII* and *nysRIII* genes (in-

- 93 -

cis), while in the latter strain *Perme** promoter is placed upstream of *nysRII* gene (*in-trans*). Nystatin production by the C3A1 and C3E1 strains was increased by 18% and 22%, respectively, compared to the wild-type *S. noursei*. Moreover, during fermentation experiments it was noticed that nystatin production by the C3E1 strain reached its maximum 24 h earlier than the wild-type strain. These data support the assumption that the *nysRII* gene encodes a positive regulator that may be used for enhancing the production of nystatin and its derivatives in *S. noursei* strains.

Expression of *nysRIV* under control of the *Perme promoter in *S. noursei* ATCC 11455:**

The start codon for *nysRIV* was reassigned, and is likely to be located 48 nt upstream of the originally proposed start nucleotide. Thus, *nysRIV* presumably encodes a 226 aa (long) rather than a 210 aa (short) protein as was previously suggested. The long and short versions of the *nysRIV* gene were PCR-amplified from the N58 recombinant phage DNA with oligonucleotide primers NR4P3 (5' - CTCAGCATGCCGAAAGGATGGCGG-3') (SEQ ID NO. 46) and NR4P5 (5' -AGGCAAGCTTCGGCGACACGGGCGT-3') (SEQ ID NO. 47), or NR4P4 (5' -CTCAGCATGCGTACGACCGGCGGG-3') (SEQ ID NO. 48) and NR4P5, respectively. The conditions for PCR were as described above. The corresponding PCR products of 0.78 kb and 0.73 kb were digested with *SphI* and *HindIII*, and ligated, together with the 0.3 kb *EcoRI-SphI* fragment containing the *Perme** promoter, with the *EcoRI-HindIII* digested pSOK804 vector, resulting in plasmids pNR4EL and pNR4ES, respectively. Both plasmids were introduced into *S. noursei* ATCC 11455 generating mutant strains NR4EL and NR4ES, respectively. Nystatin production by the strain NR4ES (expressing a 210 aa protein from the *Perme** promoter) did not differ significantly from that of the wild-type *S. noursei* harboring only pSOK804, while the NR4EL recombinant

- 93 -

cis), while in the latter strain *PermeE** promoter is placed upstream of *nysRII* gene (in-trans). Nystatin production by the C3A1 and C3E1 strains was increased by 18% and 22%, respectively, compared to the wild-type *S. noursei*. Moreover, during fermentation experiments it was noticed that nystatin production by the C3E1 strain reached its maximum 24 h earlier than the wild-type strain. These data support the assumption that the *nysRII* gene encodes a positive regulator that may be used for enhancing the production of nystatin and its derivatives in *S. noursei* strains.

Expression of *nysRIV* under control of the *PermeE promoter in *S. noursei* ATCC 11455:**

The start codon for *nysRIV* was reassigned, and is likely to be located 48 nt upstream of the originally proposed start nucleotide. Thus, *nysRIV* presumably encodes a 226 aa (long) rather than a 210 aa (short) protein as was previously suggested. The long and short versions of the *nysRIV* gene were PCR-amplified from the N58 recombinant phage DNA with oligonucleotide primers NR4P3 (5' - CTCAGCATGCCGAAAGGATGGCGG-3') (SEQ ID NO. 46) and NR4P5 (5' - AGGCAAGCTTCGGCGACACGGGCGT-3') (SEQ ID NO. 47), or NR4P4 (5' - CTCAGCATGCGTACGACCGGCGGG-3') (SEQ ID NO. 48) and NR4P5, respectively. The conditions for PCR were as described above. The corresponding PCR products of 0.78 kb and 0.73 kb were digested with *SphI* and *HindIII*, and ligated, together with the 0.3 kb *EcoRI-SphI* fragment containing the *PermeE** promoter, with the *EcoRI-HindIII* digested pSOK804 vector, resulting in plasmids pNR4EL and pNR4ES, respectively. Both plasmids were introduced into *S. noursei* ATCC 11455 generating mutant strains NR4EL and NR4ES, respectively. Nystatin production by the strain NR4ES (expressing a 210 aa protein from the *PermeE** promoter) did not differ significantly from that of the wild-type *S. noursei* harboring only pSOK804, while the NR4EL recombinant

- 94 -

(expressing a 226 aa protein from the *PermeE** promoter)
produced nystatin at a level 36 % above the wild-type
level. These data suggest that the longer, 226 aa - long
version, represents the actual NysRIV polypeptide.

- 5 Moreover, these data support the assumption that *nysRIV*
gene encodes a positive regulator that may be used for
enhancing the production of nystatin and its derivatives
in *S. noursei* strains.

- 95 -

Claims

1. A nucleic acid molecule comprising:
- (a) a nucleotide sequence as shown in SEQ ID No. 35;
- 5 or
- (b) a nucleotide sequence which is the complement of SEQ ID No. 35; or
- (c) a nucleotide sequence which is degenerate with SEQ ID No. 35; or
- 10 (d) a nucleotide sequence hybridising under conditions of high stringency to SEQ ID No. 35, to the complement of SEQ ID No. 35, or to a hybridisation probe derived from SEQ ID No. 35 or the complement thereof; or
- (e) a nucleotide sequence having at least 80% sequence
- 15 identity with SEQ ID No. 35; or
- (f) a nucleotide sequence having at least 65% sequence identity with SEQ ID No. 35 wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof.
- 20
2. A nucleic acid molecule as claimed in claim 1, said molecule comprising:
- (a) a nucleotide sequence as shown in SEQ ID No. 1 and/or in SEQ ID No. 2; or
- 25 (b) a nucleotide sequence which is the complement of SEQ ID No. 1 and/or SEQ ID No. 2; or
- (c) a nucleotide sequence which is degenerate with SEQ ID No. 1 and/or SEQ ID No. 2; or
- (d) a nucleotide sequence hybridising under conditions
- 30 of high stringency to SEQ ID No. 1 and/or SEQ ID No. 2, to the complement of SEQ ID No. 1 and/or SEQ ID No. 2, or to a hybridisation probe derived from SEQ ID Nos. 1 and/or 2 or the complements thereof; or
- (e) a nucleotide sequence having at least 65% sequence
- 35 identity with SEQ ID No. 1 and/or SEQ ID No. 2, wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part

- 96 -

thereof.

3. A nucleic acid molecule as claimed in claim 1 or claim 2 which comprises a nucleotide sequence having at least 70% sequence identity with SEQ ID No. 35, or SEQ ID No. 1 and/or SEQ ID No. 2, wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof.
4. A nucleic acid molecule comprising a part of a nucleotide sequence as defined in claim 1 or claim 2, wherein said part is at least 15 nucleotides in length.
5. A nucleic acid molecule as defined claimed in claim 1, which does not comprise the nucleotide sequence encoding ORF 1 as set out in Table 1.
6. A nucleic acid molecule as defined claimed in claim 1, which does not comprise the nucleotide sequence encoding ORF 2 as set out in Table 1.
7. A nucleic acid molecule as claimed in any one of claims 1 to 6 which encodes one or more polypeptides, or comprises one or more genetic elements, having functional activity in the synthesis of a macrolide antibiotic or a polyketide moiety.
8. A nucleic acid molecule as claimed in claim 7, wherein said macrolide antibiotic or polyketide moiety is nystatin or a nystatin derivative.
9. A nucleic acid molecule as claimed in any one of claims 1 to 6, said molecule comprising one or more genes, and/or one or more regulatory sequences, and/or one or more modules, or enzymatic domains, or non-coding or coding functional genetic elements of a PKS gene cluster.

- 97 -

10. A nucleic acid molecule as claimed in any one of claims 3 to 9, comprising a nucleotide sequence defined by any one or more of the nucleotide "start" and "end" positions within SEQ ID Nos. 35, 1 or 2, as set out in Table 1.

11. A nucleic acid molecule as claimed in any one of claims 3 to 9, comprising a nucleotide sequence defined as lying between any one or more of the nucleotide sequences having the "start" and "end" positions within SEQ ID Nos. 35, 1 or 2, as set out in Table 1.

12. A nucleic acid molecule comprising a nucleotide sequence encoding one or more amino acid sequences selected from SEQ ID Nos 3 to 20 or 36 to 43, or a nucleotide sequence which is complementary thereto or degenerate therewith or comprising a nucleotide sequence which encodes one or more amino acid sequences which exhibit at least 60% sequence identity with any one of SEQ ID Nos. 3 to 20 or 36 to 43.

13. A nucleic acid molecule as claimed in claim 12 which encodes one or more amino acid sequences which exhibit at least 85% sequence identity with any one of SEQ ID Nos. 3 to 20 or 36 to 43.

14. A polypeptide encoded by a nucleic acid molecule as defined in any one of claims 1 to 13.

15. A polypeptide as claimed in claim 14, comprising:
(a) all or part of an amino acid sequence as shown in any one or more of SEQ ID Nos. 3 to 20 or 36 to 43; or
(b) all or part of an amino acid sequence which has at least 60% sequence identity with any one or more of SEQ ID Nos. 3 to 20 or 36 to 43.

- 98 -

16. A polypeptide as claimed in claim 14, wherein said amino acid sequence at (b) has at least 85% sequence identity with any one or more of SEQ ID Nos. 3 to 20 or 36 to 43.

5

17. A polypeptide as claimed in any one of claims 14 to 16 or claim 11 having functional activity in the synthesis of a macrolide antibiotic or polyketide moiety.

10

18. A polypeptide as claimed in any one of claims 14 to 17, comprising a functional region of any one of SEQ Nos. 3 to 20 or 36 to 43, said functional region being as defined in Table 2.

15

19. An expression vector comprising a nucleic acid molecule as defined in any one of claims 1 to 13.

20

20. A host cell or transgenic organism comprising a nucleic acid molecule as defined in any one of claims 1 to 13 or an expression vector as defined in claim 19.

25

21. A method for producing a polyketide or macrolide molecule, said method comprising expressing within a host cell a nucleic acid molecule as defined in any one of claims 1 to 13.

30

22. Use of a nucleic acid molecule as defined in any one of claims 1 to 13 in the preparation of a modified PKS system, or in the preparation of a modified polyketide molecule.

35

23. A method for preparing a modified PKS system, or a modified polyketide molecule, said method comprising modifying a nucleic acid molecule as defined in any one of claims 1 to 13 or introducing a nucleic acid molecule as defined in any one of claims 1 to 13 into a

- 99 -

further PKS-encoding nucleic acid molecule.

24. A method as claimed in claim 23, being a method of preparing a nucleic acid molecule which comprises a nucleotide sequence encoding a modified polyketide synthase enzyme or enzyme system, said method comprising (a) using a nucleic acid molecule as defined in any one of claims 1 to 13 as a scaffold and modifying one or more portions of the nucleic acid molecule that encode enzymatic or other functional activities; or (b) introducing one or more portions of the nucleotide sequence that encode enzymatic or other functional activities into an alternative (i.e. different "second") PKS scaffold.

25. A method as claimed in claim 23, being a method of preparing a nucleic acid molecule comprising a nucleotide sequence encoding a modified nystatin PKS, wherein said modified nystatin PKS is derived from a nystatin PKS encoded by a nucleic acid molecule as defined in any one of claims 1 to 13, said nucleic acid molecule containing first regions which encode enzymatic or other functional activities and second regions which encode scaffolding amino acid sequences, said method comprising:
(a) modifying at least one said first region; or
(b) incorporating at least one said first region into a scaffolding-encoding second region from a different PKS-encoding nucleotide sequence.

26. A method of preparing a modified nystatin PKS as defined in claim 25, said method comprising expressing a nucleic acid molecule prepared as defined in claim 25 within a host cell under conditions whereby the modified nystatin PKS is expressed.

- 100 -

27. A modified PKS system derived from a nucleic acid molecule as defined in any one of claims 1 to 13 by a method as defined in any one of claims 23 to 26.

5 28. A modified PKS system as claimed in claim 27, comprising a modified nystatin PKS wherein one or more of the polypeptides of SEQ ID Nos. 3 to 20 or 36 to 43, or one or more of the enzymatic domains in one or more modules of SEQ ID Nos. 5 to 7, 20, or 37 to 39 are
10 modified.

29. A modified PKS system as claimed in claim 27, comprising a modified nystatin PKS wherein
(i) the ER domain in module 5 of Nys C (amino
15 acids 4953-5239 of SEQ ID No. 7) is inactivated; or
(ii) a module of Nys C (SEQ ID No. 7) is deleted;
or
(iii) the DH domain in module 8 of Nys C (amino
acids 10086-10289 of SEQ ID No. 7) is inactivated; or
20 (iv) the KR domain in module 7 of Nys C (amino
acids 8812-9086 of SEQ ID No. 7) is inactivated.

30. A multiplicity of cell colonies comprising a library of colonies wherein each colony of the library
25 contains a different modified PKS as defined in any one of claims 27 to 29.

31. A method of producing a library of PKS complexes or a library of polyketide molecules, said method
30 comprising culturing a library of colonies as defined in claim 30.

32. A library of PKS complexes or a library of polyketide molecules obtained or obtainable by a method
35 as defined in claim 31.

- 101 -

33. A combinatorial library of polyketides wherein the polyketide members of the library are synthesized by a modified PKS system as defined in any one of claims 27 to 29.

5

34. A host cell or transgenic organism containing (a) a nucleic acid molecule prepared by a method as defined in any one of claims 23 to 25 or (b) a modified PKS system as defined in any one of claims 27 to 29.

10

35. A polyketide produced or producible by a host cell or transgenic organism as defined in claim 34.

15

36. An antibiotic derived from a polyketide as defined in claim 35.

1 / 12

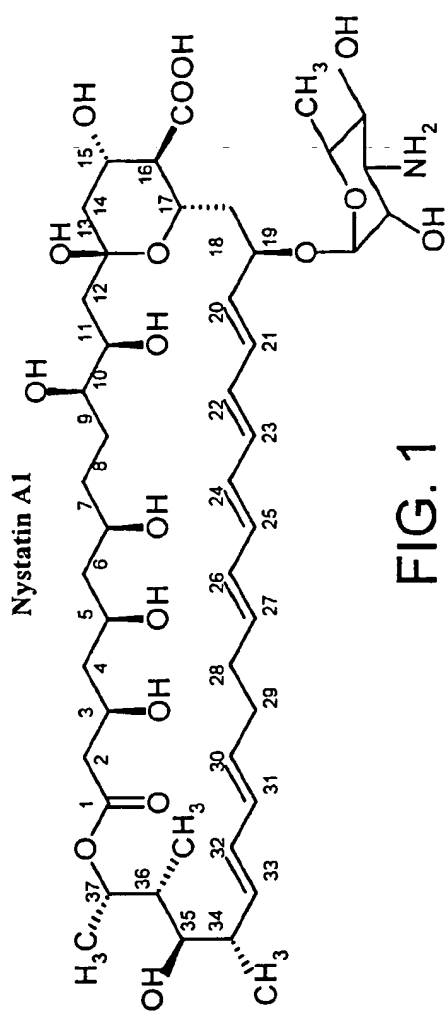


FIG. 1

2 / 12

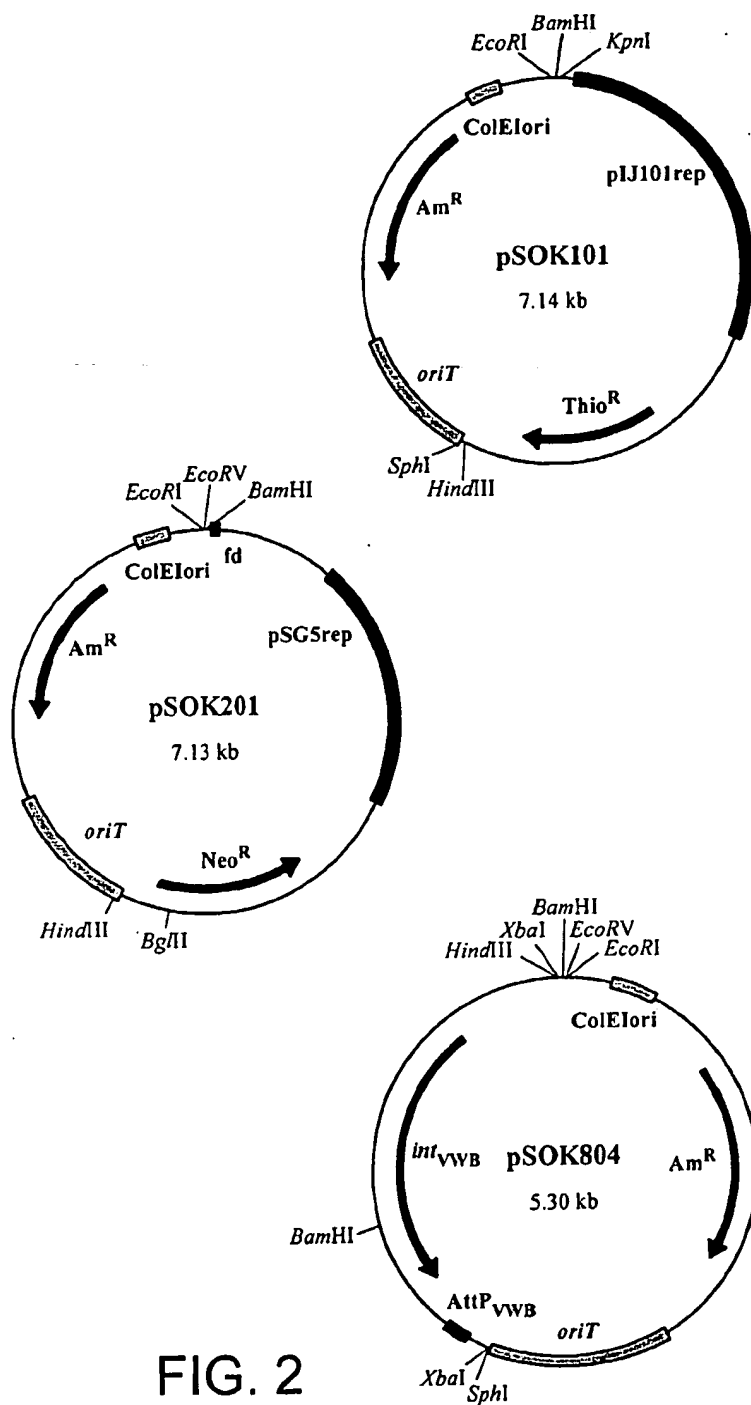


FIG. 2

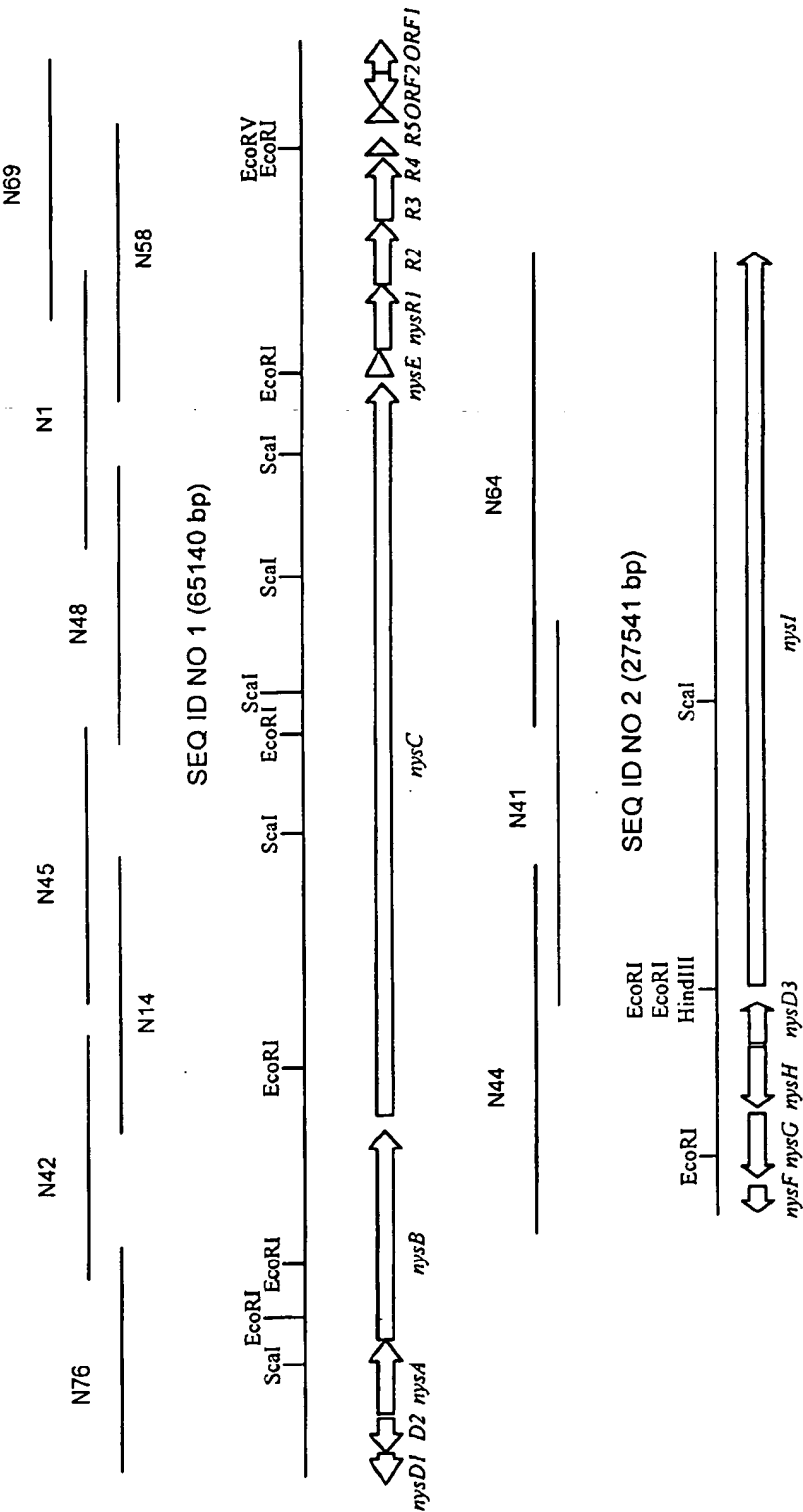


FIG. 3

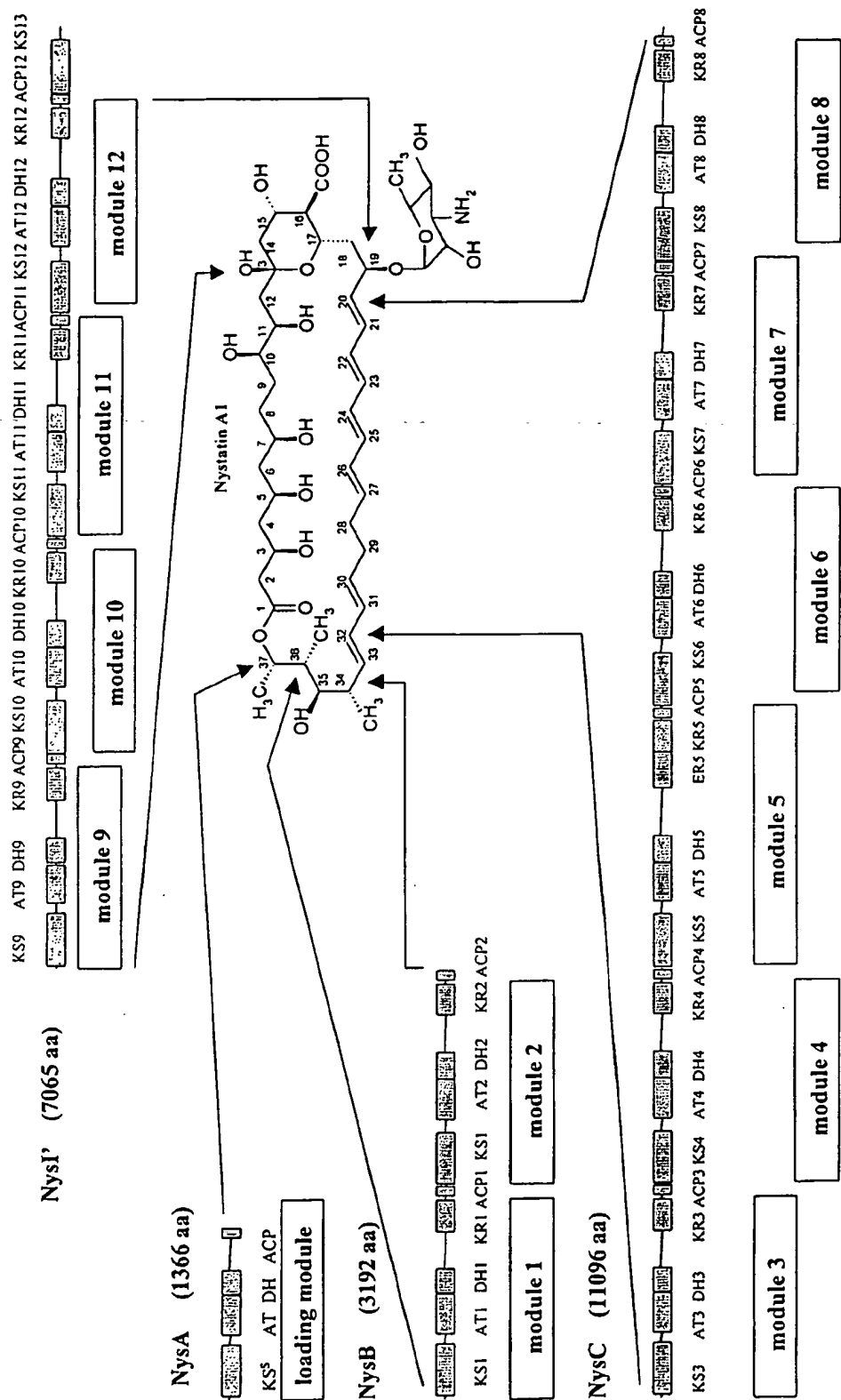
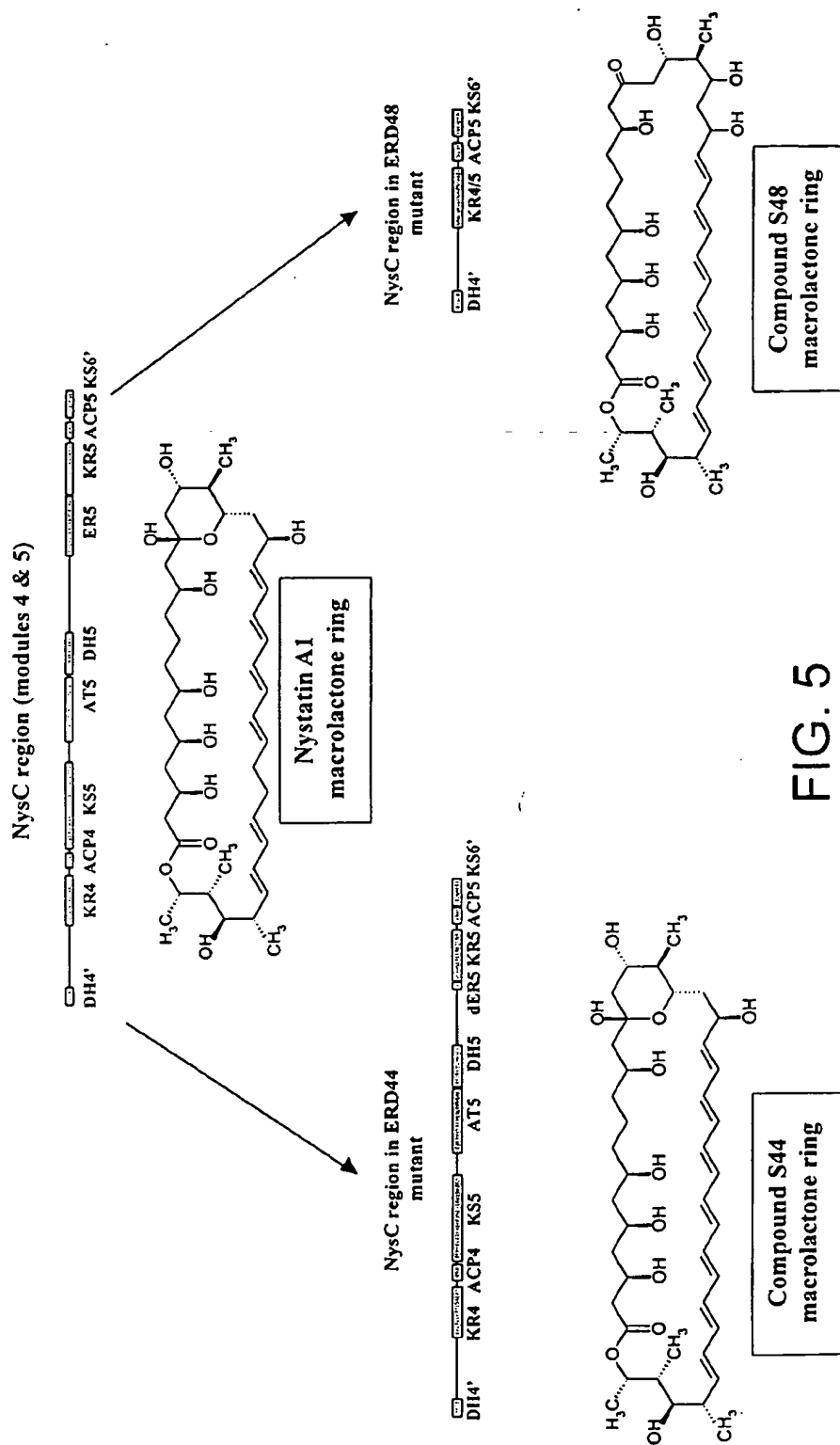


FIG. 4

5 / 12



6 / 12

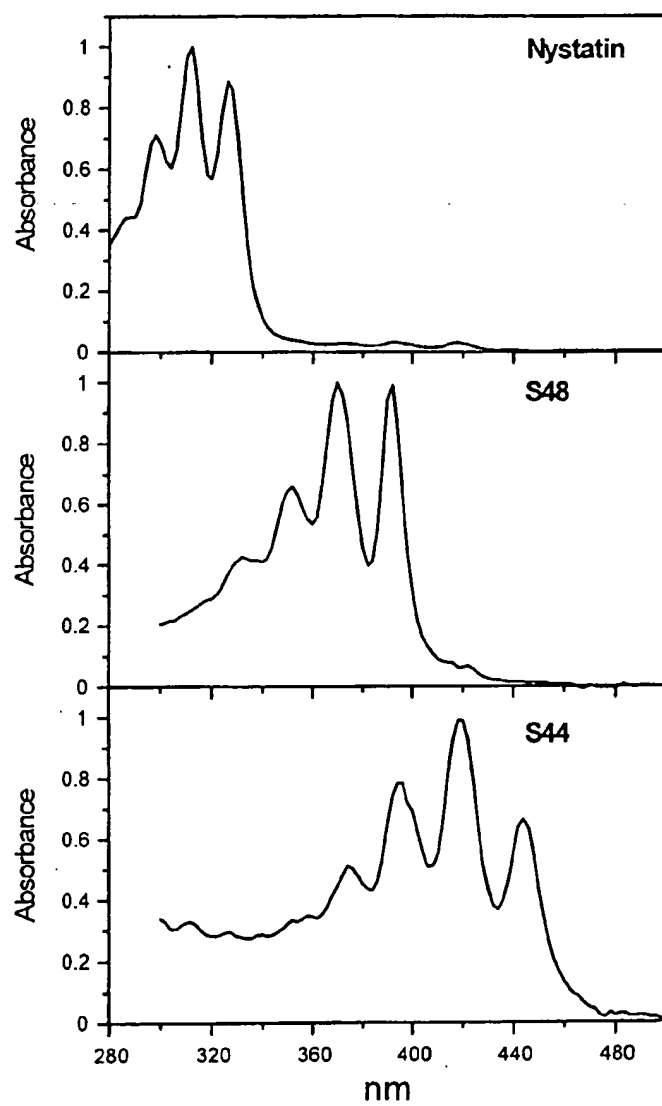


FIG. 6

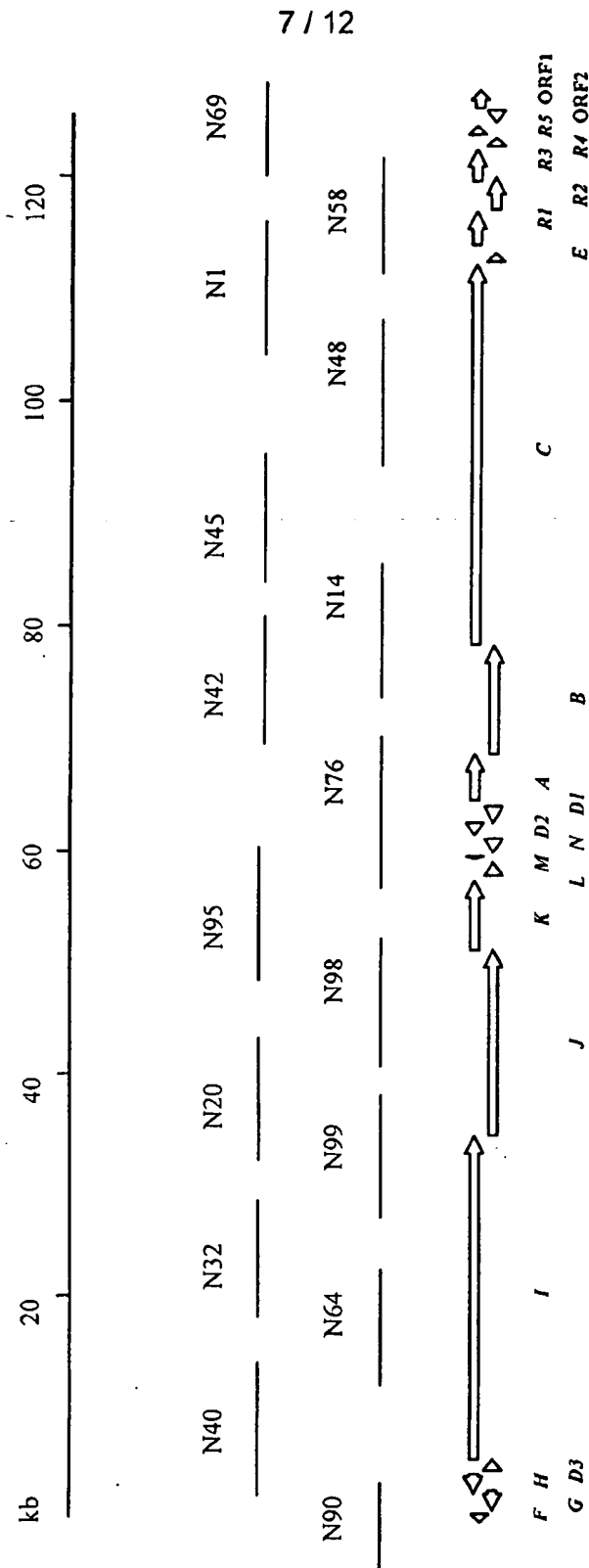


FIG. 7

8 / 12

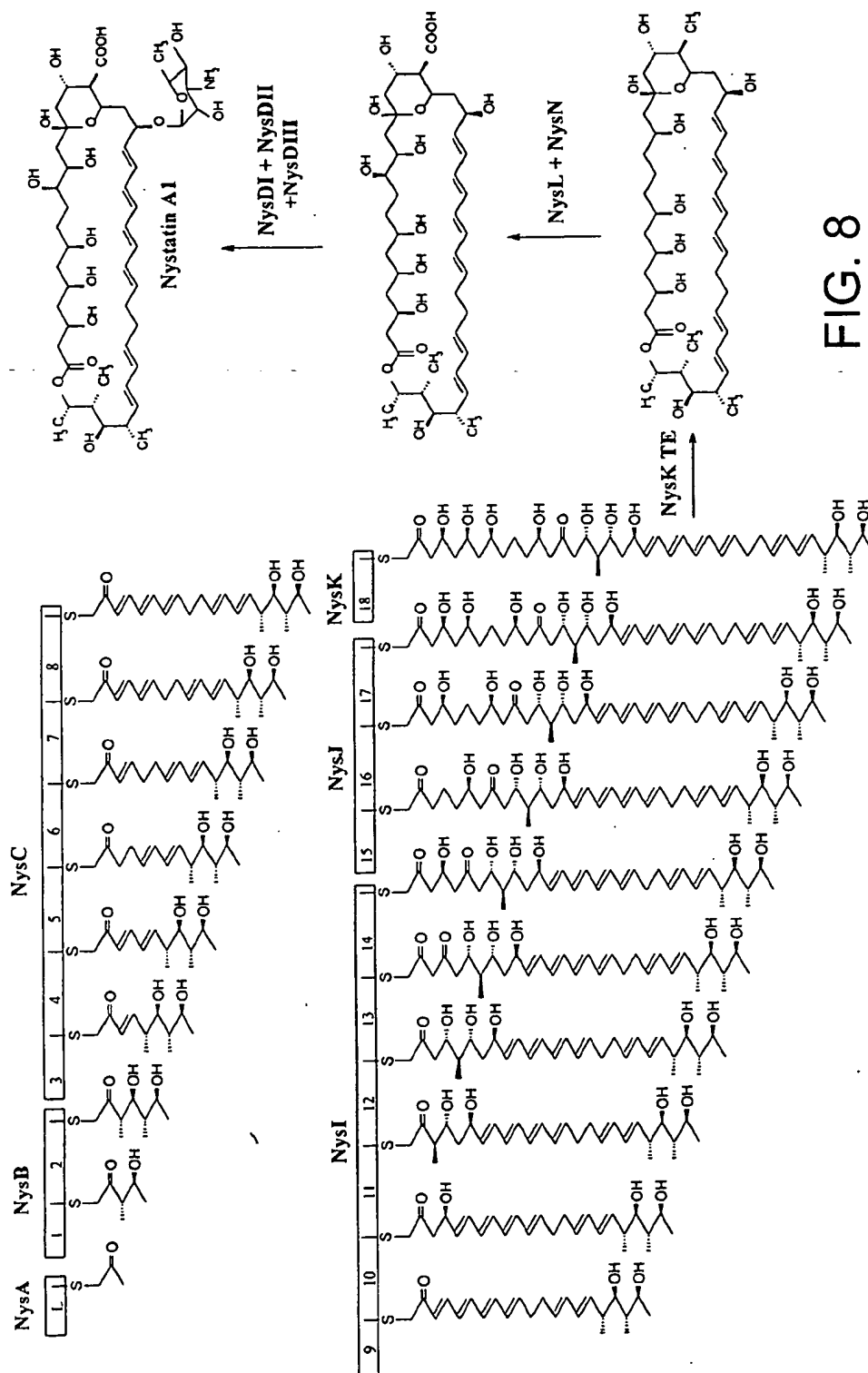


FIG. 8

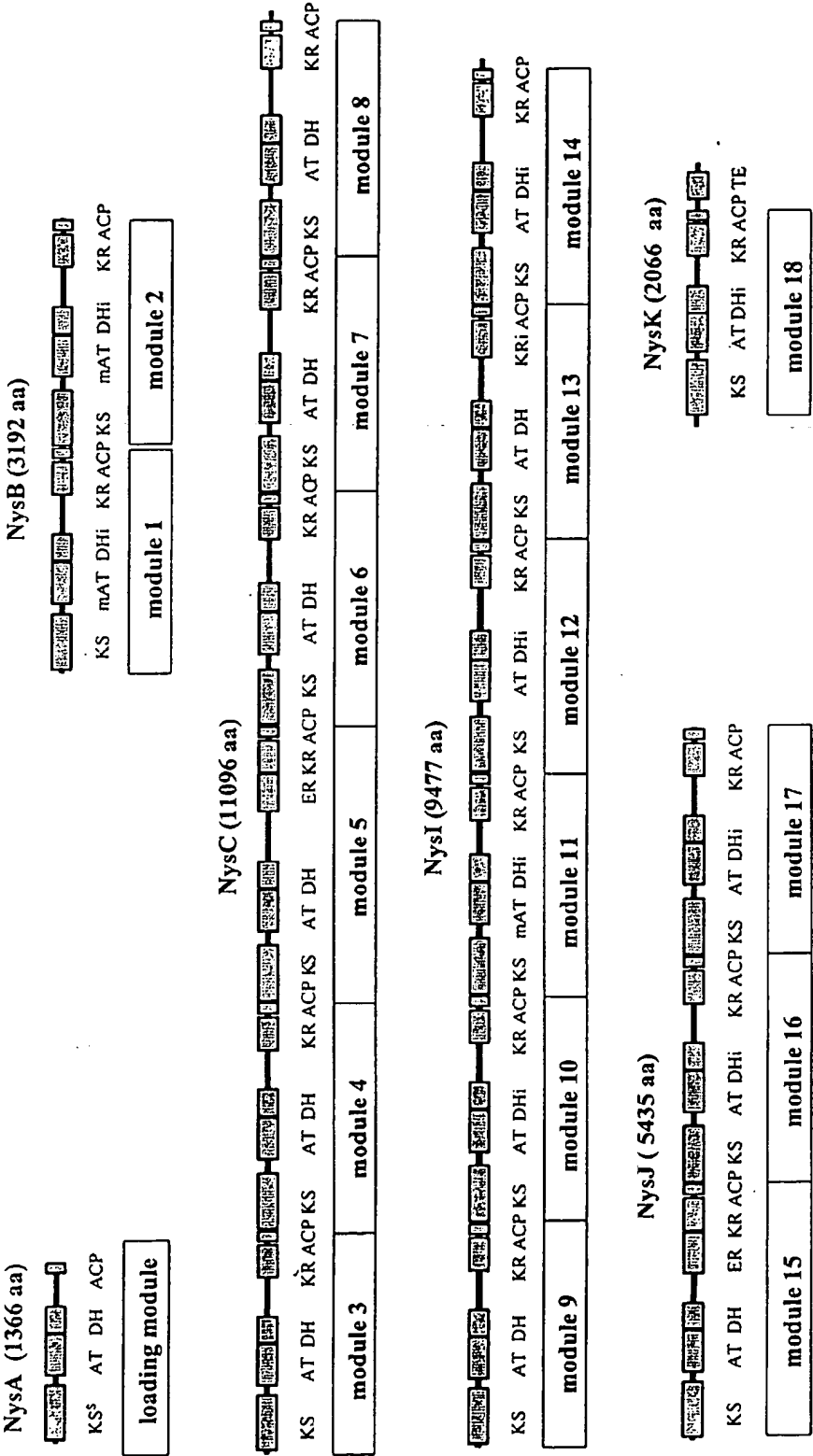


FIG. 9

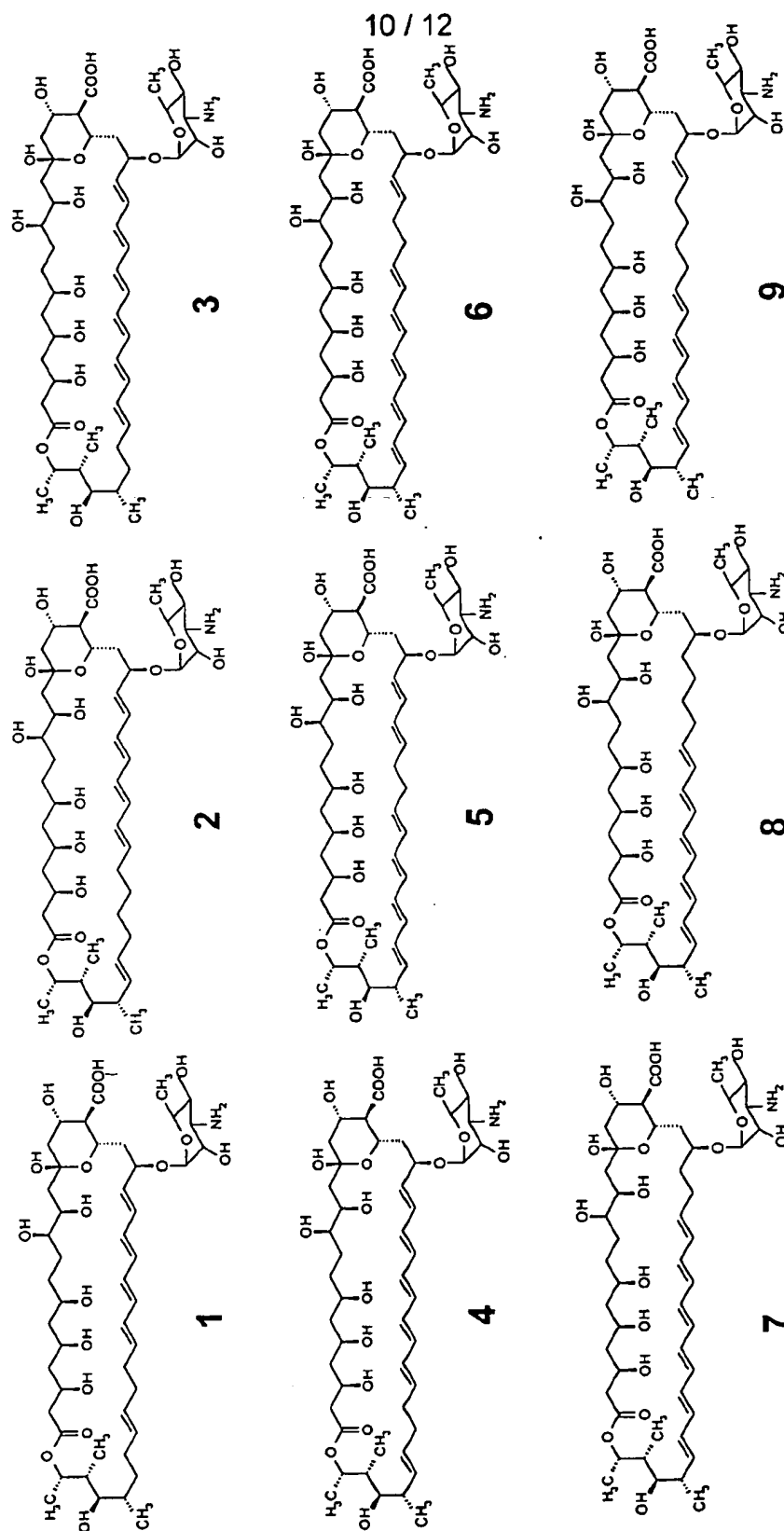
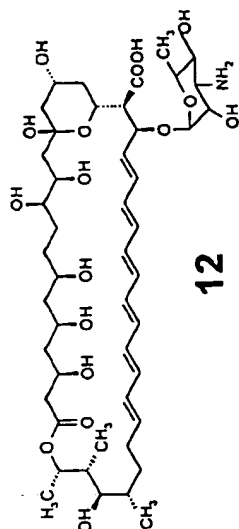
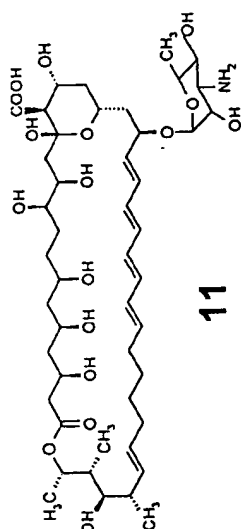


FIG. 10

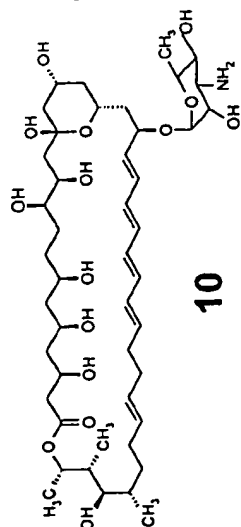
11 / 12



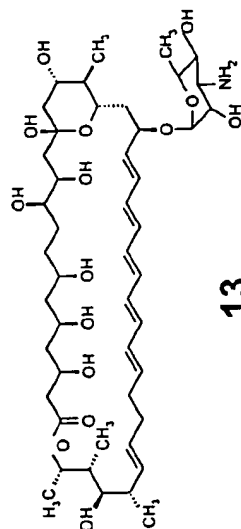
12



11



10



13

FIG. 11

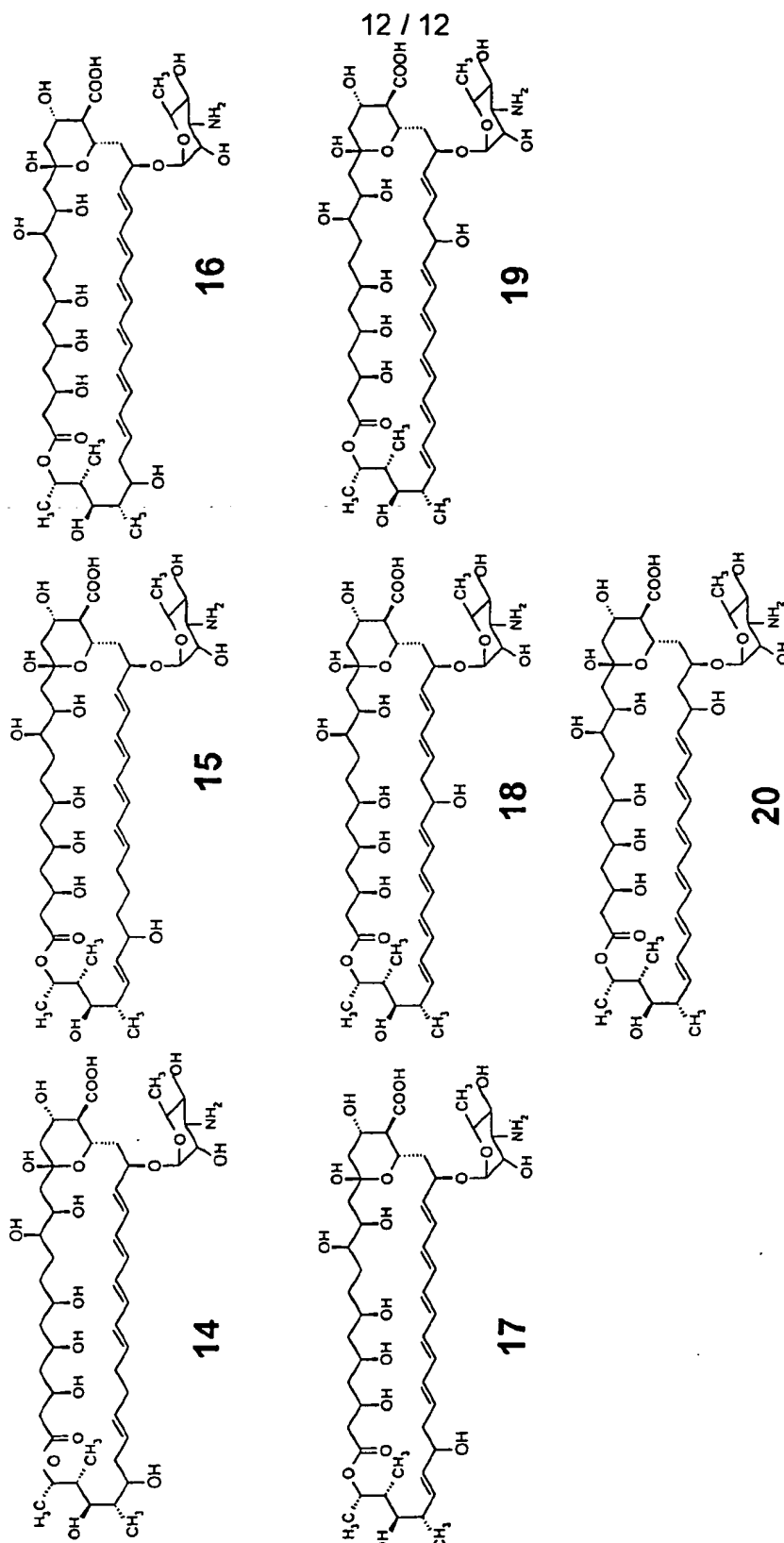


FIG. 12

1

SEQUENCE LISTING

SEQ ID NO. 1

5 DNA sequence: (Nysl; region 1)

1 gatccggccg gtgatgaact cctgctcctg tgcggtcagt ccggtgtgcg
 51 tcggcagata gaagccgtcg gcgctgagcc gggccgcgtt caacgcgggc
 101 cagtcggcgg agaagtaacc gggctgacgg ctcacgcgct tgaagaagac
 10 151 ccgggtctcg atgcctcgc cggcgaggta cgcgcacagc tcgtcccgc
 201 gtcggcccg caggctgtac atccacagca cgtcgcgcgg cggcatcagc
 251 gtgatgccgg ggatgtcgcg cagcgcctcg tcgtagcgt tctcgatgtc
 301 gcggcgagg gcgaggatgg tgtccagctg ctcggctctgc gcgagcgcca
 351 ccgggcctg catgttggtc atccggaagt ttagggccag cttcttgtgc
 15 401 aggaagctgt ggtccttggg gaacgccatc gccgcagat gggccatctg
 451 ctcggccagg tgcgggtcgt gggtcaggca gacgccgccc tcgcccggcg
 501 agatgatctt gttggcgaag agcgagaaac aggcgatgtc gccccgcggc
 551 cgcaccccggt gcgcctcggc ggagtcctcc accaccgca ggttgtaactc
 601 gtacgccagg ttcagcacgg cgtecatgtc gactgcccgg ccgtagatgt
 20 651 gcaccggcat gatcactttg gtgcgcgggg tgatcttctc ctcgatgcgg
 701 gacacgtcga tgttcagggt gtcgccgcag tccacgaaca ccggcggtggc
 751 accggtgtag gtcaccgccc aggcggagcg gatcatcgtg aactccggga
 801 cgatcacctc gtcaccgggg ccgacgcccga gcgcgcgcag cgcacgcgtc
 851 agcgcctggtg tgccggagga gcaggcgacg ccgaacggca cgtcgttgta
 25 901 cgcggcgaac gcctcctcga accgcctgac gtacggcccc tgcgaagaga
 951 tccagccgcc gccgacggcc tccgtcacat agtcgagctc gcggccctgg
 1001 agccacggca tggacaccgg atacgtaaag gacatgggtt ttgagtcttc
 1051 ctcggtcagt cggttgccag ggcgggaagg ccgagcagca ggtccgcgc
 1101 cgcggcccgg ccgccggcgt cccgcagcag cccggcgaag tgctccgcgc
 30 1151 gtcggtgaa ggagggttg tgcagcacgc gggatgatctt gtccaggacg
 1201 tcctcgggtgt ccacggtctc cggccgggtcc agggtcaggc tcaccccgaa
 1251 gtcctggccc cggatcgctt ggtcgtcgca gtccaccac aacggccgga
 1301 ccaccagcgg ctttccgaag tacaggccct cgtggtagcc gttgccaccg
 1351 gcatgggtga agaacgcctt cacgttcgga tgggccagca cgtccagctg
 35 1401 cgacggcacc cagccctcga tccgcagggt gtccggcagc tcggcgcccg
 1451 gcggcagcaa ctctgttg cgcgcgggga gtttccacaa cacctggtgg
 1501 ccccgccgt ccagtcgccc ggcgacctcc accagcgacg ccacctgctc

1551 acgggtcagc cgggtgatcg tgccgaagcc catgtacacc acggacttct
 1601 gcgccgacag ccagtccgac aggccgtcgt cgtccggtgc ctggggcagc
 1651 ggcggcacca tcgtgcccac cagccgcagc ttcggatgca tcgggaacgg
 1701 gtagtccaac tcccttacgg agtagcacia gacctgctcc gcatgggtcga
 5 1751 tccgcgccat catctgccgc gcctggggcg cgatgcccag ctcggtgcgg
 1801 acccggttgt cctcctcgac gaccttgccg acgtccgacg tcaggaacat
 1851 cccgagcgtc cgcagccgga acagctggtt ctcgatccgc tgagccaggg
 1901 acatcgccgc cggcagcccc gagtgccgca ccgggaaacc cgacggggtg
 1951 taggacttgg cgaacgggac gtgagaggtg aggacgttgc tcggcacgaa
 10 2001 cggcaccgcc agcacgaacg gaatgccctt ggtgatcgcc agctcgtacc
 2051 cgaactggca catgctctcg atcaccatca gcgccggctc gacctcctcg
 2101 acgatctcct ccaggcggcg gtacttcgcc atccgcgact ccggcgcgaa
 2151 cgaatgccga atcaccgccg cgtgcgcctt gaaccgcgac cgctgcgtca
 2201 cctccgcata cgtcgcgctc tcccatgtga ccgccgacat ctgcgagacg
 15 2251 gtgtcgccga gcgacgcgaa ccgaaccggg ctgccgtcca ccacggccgc
 2301 cacctcgctc cgcgctttct cgtcgggtggc gaaccacagg tccgccacgt
 2351 cgcgcgggga caattccccg gccagcacga gcagcggatt gagcaggccg
 2401 ctttcggcat aactgacgaa caggatcggc cgccgattcg cgcccatgga
 2451 caacacccct cggaaatgtg cgggccgccg ggcccgcgcg ccacgcaccc
 20 2501 gcccggcccc gtcgcggggt gagtgcatc gccgacgccg ccacccgagg
 2551 cgcgtgttgc cggaaaggaag ggtcaccggc cggcaccgcc aacgcgccgc
 2601 gtggaaaacg ggtcggttac ttggtctcat gccacggacc ggggaatcac
 2651 tagtcttcgg cgcgcgacgg ccctttcccg gccgtgtggc caatgcccgt
 2701 ccccggcgcc cgtcattcct tagggaaaag tacagcgttt gcgaacgtac
 25 2751 gatccggcac gcagaggtga cctgaggcca acttttccgc aggggtgagc
 2801 aaggcatgac gatcggagcc gacgaggacc cgggtggtgt cgtcggaatg
 2851 gcctgcggtt atccgggtgg ggtcgcggc ccggaggacc tgtgggaact
 2901 ggtccgcacc ggccgcgacg cgaccaccgc cttcccggac gaccgcggct
 2951 gggacctggc cgcactggcc ggcgacggac ccggccgcag cgcgaccgc
 30 3001 gagggcggat tcctcaccgg cgccgccgac ttcgacgccg cttcttcgg
 3051 catgtcgccc cgcgaggccg tctccaccga cccgcaacag cgctcgtcc
 3101 tggagaccgc ctgggaagcc ctggagcgcg ccggcatcga cccgcactcc
 3151 ctgcgcggca gccgcaccgg ggtcttcgtc ggcgccagcg gccaggacta
 3201 cgccgccgtc acccacgcct cgcccgacga cctggacgga cagccctca
 35 3251 ccggcctggc ccccgccgct gcctccggtc gcctggcgta cgtcctgggc
 3301 ctcgaaggcc ccgccgtcac cgtcgacacc acgtcctcct cgtcgtgggt
 3351 cgcgctgcac tgggcgggtc gcgccctgcg cgcgggggag tgcagcaccg

3401 ccctggccgg cggcgtcacg gtgatgtcca cccgggccgc cttcgtcggc
 3451 cacacccgac agggcggcct cgcgcccgac ggccgctgca agccgttctc
 3501 cgacgacgcc gacggcaccg cctggggcga gggcgctggc atcgtcgtcc
 3551 tggagcacct gtccaccgcc cgcgcgcggc gcaaccccg cctcgcgtg
 5 3601 ctgcgcggct cggccgtcaa ccaggacggc gcctccgacg gcctcaccgc
 3651 acccagcggc cccgcccagg aacgcgtcat ccgcgccgcc ctgcgcgacg
 3701 cccgactcgc ccccgccgac atcgatctcg tcgaggcgca cggcaccggc
 3751 acccggctcg gcgaccccg cgaggcccg gcgctgctcg ccgcctacgg
 3801 ccaggaccgg gaccgggacc gaccgctgcg cctcgggttc ctgaagtcca
 10 3851 ccctcggcca cgcacaggcc gccgccggca tcggcggagt gatcaagacc
 3901 gtcctgaccc tgcggcacgg cctgatgccg cgcacccggc acctggccac
 3951 cccacccgc caagtgcact ggtcccagg cgccgtggcc cccctcaccg
 4001 accacacgcc ctggccaccg gccgaccgac cgcgcgcgcg cggcgtctcc
 4051 tccttcggca tcagcggcac caacgcccac gtgatcctcg aagaggcgcc
 15 4101 gcccgcggac gtcccgtca cccggcccg caccctcgc cccagcaccg
 4151 tcccctggcc ggtctccgcc gccacggcc aagccctcga cgcaccaactc
 4201 gcccggctcc gcgccaccct gcgcaccac tcggacctgg acccgtgga
 4251 cgtcggctac tccctggcca ccggccgcgc cgcgtccgc caccggcg
 4301 tcctcctgcc gcccgcgac ggacccgcg cggacgccgt cgagcacgcc
 20 4351 cgcggtgcgg cccaccagcg ccgcaccgcc gtccctcttct ccggccaggg
 4401 cagccagcgc ccgggcatgg gccgcgaact cgcgcgccgc tcccgggtgt
 4451 tcgccgacgc actggacgac gcgctgcgcg ccctggaccg gcacctggac
 4501 ggcccggtgc gcgaggtgat gtggggcacc gacgccgcgc tcctggaccg
 4551 gaccggctgg acccagcccg cctgttcgc cgtcgaggtc gccctccacc
 25 4601 gcctggtcgc gtccctcggc gtcaccccg acttcgtcgg cggccactcc
 4651 gtcggcgaga tcgccgcgc ccacgtcgc ggcgctccta gcctggagga
 4701 cgcctgccgc ctggtggcgg ccgcgccac gctgatgcag gcgctccgg
 4751 ccggcggcgc gatggccgc ctggaggcca ccgaggacga agtggccccg
 4801 ctgctcggcg cacacctgc gctggccgcc gtcaacggcc ccaccgggt
 30 4851 cgtcgtcgcc ggagccgagg acgccgtgc gcaactgacc gcccgcttcg
 4901 ccgaccgcgg ccggcgacc agccggctgg ccgtctcgca cgccttcac
 4951 tcgccgctga tggagcccat gctcgacgcc ttccgggacg tcgtgagccg
 5001 actgaccttc caccagccgt cgatcccgct ggtctccaac ctcaccgggtg
 5051 aactcgccgg cagtgcgac accagcggc agtactgggt ccggcacgctc
 35 5101 cgcgacaccg tccgcttcgc cgacggcatc accgcactgg ccaaggccgg
 5151 cggcgacgct ctgatcgaac tcggccccgg cggcgtgctg tccgcgatgg
 5201 cccgcgacac cctcggcccc gacagacca ccgacgtcgt ccccgccctg

5251 agcaaggagac ggccccgagga gaccgccttc gccggcgccc tcggccgcct
 5301 gcacaccctc ggcgtccccg tcgactggcc cgccttctac gccggcaccg
 5351 gcgcccgcgc cgtcgaactg cccacctacg ctttcagca cgtgcgccac
 5401 tggccccacc cgcggcgccc gaacggcgcc gggcccgcg ccctcggcc
 5 5451 cccctgctc ggctccgcgc tcgaactcgc cgacggcggc ggcaccgtct
 5501 gctccggcgc cctctccctc cgcacccacc cctggctcgc cgaccacacc
 5551 gtcgcccggc gggtcgtgct gccggccacc gcgctgctgg aactcgccgt
 5601 gcgcccggc gacgaggcgc gctgcgacgt cctgcacgaa ctccacctca
 5651 ccacccgcgc ggccctgccc gacgacgccg ccctgcacgt ccagggtcac
 10 5701 gtcggccccg ccgacaccac cgggcgcgc gccgtcaccg tccacaccg
 5751 ccccgaccac caccggcgc gcgactggac ccgatgcgc accggcacc
 5801 tcggcagcac ccgcgcgtcc gcagccgaag ccgccacggg cggcaccgc
 5851 gccgcctggc cgcggcgcca cgcgaaccc ctgcacctc cgcaccacta
 5901 cgaacggctc gccgacgcgc gcttcgacta cggcccgacc ttccgcggcc
 15 5951 tgcggggcgc ctggcgacgc ggcgcggaga tcttcgcgga cgtggaatgc
 6001 ccgcccggca ccgcccacga cggcccgac cagggactgc accccgcct
 6051 gctcgacgc gcccggcac ccgccatggc ggtggacggc accgtgccc
 6101 tcgcctggca cggcgtccgc ctgcacgcgc tcggcgccac cgcgctgcgc
 6151 gtccgcatcc gcccaccac gaccggcacg ctgacctca ccgcggtcga
 20 6201 cgtgcacggc gcgcggctc tcaccgtcga ggccctcacc gcccgccgc
 6251 tgaccgacga ggaacgcgc gcccgcgga cgcgcggca ggcgcggc
 6301 gagacgccc ccgacgccc ccggcccg cgcgcggc cccgccccg
 6351 cccggcggc gaacccctc cggacaccac cgggtccac cccaccgc
 6401 gccacctcgc cgcgctgcgc ccggccgccc gggagcgcca gctgctggac
 25 6451 ctggtgcga cccaggcgc cgcgtcctg ggccacccg gccccgaggc
 6501 cgtcggcacc cgcagcgtc tcaaggagc gggcttcgac tcgttgccg
 6551 gcgtcgaact cgcgcaccg ctcaccgcc gcaccggact gcgcctgcgc
 6601 gccaccctc tcttcaact cccaccccc gaacgtgctg cccaccgcct
 6651 cggggaactc ctgcgcgca cggccccc cgcacccggg gcgtacggag
 30 6701 aggaactcac caggttcgag gcgatcgtga cgaacctgcc gcaggacggc
 6751 cccgaacgcc gggccgtcgc ggaccggtg gacgccatc tctccgact
 6801 ccgcccagaa tcgcctgcag aggtgccctc ctcggacgag gacatcgaca
 6851 cgggtgcggt cgacagactg ctgcacatca tcgatgaaga gttcgaaacc
 6901 acatagagaa attgttgctt tcgttcgca cccgatgac aggacggacc
 35 6951 gatgcaggaa cccagcaag gccagccgga ccagcaggag aaaatcgctc
 7001 actatctcc gcgggtcact tcagatcttc gccgtgccc cgcgcgact
 7051 ggcgaactgg aatccaagga caacgagccc atcgccatc tcggaatggg

7101 ctgccgactt cccggcgggc tcaattcgcc ggaatccctg tgggacctgg
 7151 tgcgttccgg cggcgacgcc atttccggat tccccgtcga ccgcggtcgg
 7201 gacctgaaa ccctcaccgg aaacggcgac ggcagcagcg ccacccacga
 7251 aggcggattc ctctacgacg ccgcggaatt cgacgcgcgc ttcttcggca
 5 7301 tctcgccgcg cgaggcgact gctatggacc cccagcagcg cctcctcctc
 7351 gaagtcgcct gggaggcgct ggagcgcgcc ggcacgccc ccacagccct
 7401 gcgcggcagc cggtcggcg tggttcgtcgg ctctaccac tggggcgcg
 7451 cctcggccga cgccgccacc gaactgcacg gccacgccct gaccggcacc
 7501 gccgccagcg tgctgtccgg ccgectggcc tacaccctcg gcctcgaagg
 10 7551 cccggccgtc accgtcgaca ccgectgctc ctctccctc gtgcctcgc
 7601 acctggcggc ccagtcctcg cgcgtcggcg aatcctcgct cgccgtgatc
 7651 ggcggcgctc cgtacctcac cgagccgtcc gtcttcgtcg agttcagcgc
 7701 ccaggcgggc ctggcaccgg acggccgctg caaggcgctc tccgacgccg
 7751 ccgacggcac cggttgggccc gagggcgctg gtgtcctcgt cgccgagcgg
 15 7801 ctctccgacg cgacgcgcaa cggccatccg gtcctcgccg tgctgcgcgg
 7851 ctccgccgtc aaccaggacg gcgcctccaa cggcctgacc gcccccaacg
 7901 gcccctccca ggaacgggtc atccagcagg cctcgcgccg gaccggcctg
 7951 acccccgccc acatcgacgc cgctcgaggcg caccggcaccg gaccccggt
 8001 cggcgacccc atcgaggccc aggcctgct cgccacctac ggccagggac
 20 8051 acacccccga ccagccgctg tggctcggct cctgaagtc caacatcggg
 8101 cacacccagg cggccgcggc cgctcgccggt gtcataaaga tggatcatggc
 8151 gctgcgccac ggccacctgc cgccgacct gcacgcgcac gcgcctcct
 8201 cgacagtgga ctggtccgccc ggatcggtac gctgtgtgac cgagggccag
 8251 cagtggccgg agaccggacg tccgcgcggc gccgcggtgt cctcgctcgg
 25 8301 catcagcggc accaacgcgc acgcctgct ggaacaggca cccaccccg
 8351 cggacaccgc ggacgcgggc gacgacgcgc cgcccaccga accggccggc
 8401 gcgcccgcgc cgctgcctg gatcgctctc ggacactccc cgcaggcgct
 8451 gcgcgaccag gccgcgcgcc tggccgccag ggtcgagacc gaccccgcg
 8501 tccgccccca ggacatcggg cacaccctgc acaccgccc cgccctgctc
 30 8551 gaacgacgcg ccgtcgctcg cgccccgcac cgcgccgaac tctcgcggc
 8601 taccacagag ttggccgccc gccggtccgc gaacgcgcgc gtcgagggcc
 8651 tcgcggacgt cgagggtcgg acggtgttcg tggtccccgg tcagggttcg
 8701 cagtgggtgg ggatgggggc ccaactcctc gatgagtcgg cgggtgttcg
 8751 ggagcggatt gccgagtggt cggcggcact cgccgagttc accgactggt
 35 8801 cgctggtcga tgtgtgcgg ggtgtggtgg gtgcgccgtc gttggagcgg
 8851 gtcgatgtgg tgcagccggc gtcgttcgcg gtgatggtgt cgttggtcgc
 8901 gttgtggggc tcccggtgtg tggtgccgga tgcggtggtg gggcattcgc

	8951	agggtgagat	cgtgcccgcg	gtggtgtcgg	gtgcgctgtc	gttgcgggac
	9001	ggggcgcggg	tggtaggcgt	gcggagtcag	gccattggtc	gtgcgttggc
	9051	ggggcggggc	gggatgatgt	ccgtcgcgct	gtcggtaggac	gtgctcgaac
	9101	cgcggttggg	cgagttcgag	gggcgggtgt	cggtggccgc	cgtcaacggc
5	9151	ccgcgctccg	tcgtggtcgc	cggcgagccc	gaggcgctgg	acgcgctgca
	9201	cgcccggtg	accgccgacg	acatccgggc	ccgccggatc	gcggtaggact
	9251	acgcctcgca	ctcgcaccag	gtcgaggacc	tgcacgagga	actgctggag
	9301	gtgctggcgg	agctggcgcc	gcgcacgtcg	gaggtgccgt	tcttctcgac
	9351	cgtgaccggc	gactggctgg	acaccgcgcg	gatggacgcc	ggctactggt
10	9401	tccgcaacct	gcgcggacgg	gtgcggttcg	cggacgcggg	ggcggacctg
	9451	ctggcgcgcg	agtaccgcgc	gttcgtcgag	gtcagctcgc	accgggtgct
	9501	gacgatggcg	gtcttggaac	tgatcgagga	ggccggggtc	acggccgtcg
	9551	cgaccggcac	cctgcgcctg	gaccaggggtg	gcgcggggccg	cttctgtctg
	9601	tcggccgccc	aggtcttcgt	gcgcgggtgtg	gacgtggact	gggcgggggc
15	9651	gttcgagggg	accgggtcgg	cccgggtcga	cctgcccacc	tacgccttcc
	9701	agcgcgagcg	gtactggaac	accgcaccg	ccgccgaccg	caccccggcc
	9751	gacgccccga	tggacgccga	attctggggc	gccgtcgaac	aggcggacgt
	9801	ctccgcgctg	accgccgcgc	tcggcaccga	cgaggactcc	gtcgccgccca
	9851	tctgcccggg	cctcacctcc	tggcgccggg	cccgtccca	gcgcaccacc
20	9901	ctcgactcct	ggcgctaccg	cgtcacctgg	acgcccctcg	cccagggtgcc
	9951	ccgcgccacc	ctgaccggca	cctgggtgct	ggtcaccacc	gacggcatcg
	10001	acgacaccga	tgtggcaggg	gcgttgagga	gctacggcgc	cgaggtagcgc
	10051	cggctggtcc	tggacgagga	gtgcaccgac	cgcgccgtcc	tgcgggagcg
	10101	gctggccggc	gcggaggacg	tgaccggcat	cgtctccgtc	ctcgcccgccg
25	10151	ccgaggacga	cgccgcacgc	caccccggcc	tcacccgggg	actcgcgctc
	10201	accgtctccc	tcgtccaggc	cctggggcgac	gccgaggcga	ccgcgccgct
	10251	gtggttctctg	accgcggcg	ccttcgccac	cggcccgtcc	gaccccgta
	10301	cccggcccct	gcagagccag	atcgcgggcg	tcggctggac	caccgcgctg
	10351	gagcaccgcg	agcgtgggg	cggcaccgtg	gacctgcccg	acaccctcga
30	10401	cgcccggggc	gcccagcggc	tcgccgcgcg	gctgtccggc	gccctcggcg
	10451	ccgaggacca	gctcgccgtc	cgcgccgccg	gggtactggc	ccgccgcac
	10501	gtgcgtgccg	gacaccgcgc	cggacgaccg	gcacggacct	gggcgccgcg
	10551	cggcaccacc	ctgatcaccg	gcggctccgg	cacctcgcc	ccgcagctcg
	10601	cccgtgggt	ggccgaacgc	ggcgccgagc	acgtgggtgt	ggtcagccgg
35	10651	cgcggtgccg	acgcccccg	agcgcccgaa	ctcatcgcg	aggcagccga
	10701	gtcgggcacc	gaggtgaccg	tcgccgcctg	cgacatcacc	gaccgcgacg
	10751	cggtcgccgc	gctgctggcc	gacctcacgg	ccgacggccg	cacctgcgc

10801 accgtcatcc acgccgccgc cgccatcgag ctgtccgcgc tcgccgacac
10851 caccgtggcg gagttcgccg acgtcgtgca cgccaaggtc accggcgcac
10901 ggatcctcga cgaactgctc gacgacgcgg aactggacga cttcgtcctg
10951 tactcctcca ccgccggcat gtggggcagc ggcgtgcacg ccgcctacgt
5 11001 cgccggcaac gcctatctgt ccgcgctcgc cgagcagcgc cgcgcccgcg
11051 gactgcgcac cacctccatc cactggggca agtggcccga cgaccgggca
11101 cgcgagctgg ccgaccgcga ccggatccgc cgcagcggtc tggagtacct
11151 cgaccccagag ctggcgctca ccgcgctcca gcacgtcctg gacgacgacg
11201 agaccgtcat cggcctcatg gacatcgact gggacaccta ccacgacgtg
10 11251 ttcaccgcgg gccggcccgc gcacctcttc gaccagatcc ccgaggtgcg
11301 gcgccgcctc gaccaggcat ccgtcccga ccccgcgggc ccggccgcgg
11351 acggcctggc cgcccgcctg caggccctcg ccgccgccga acaggaccgg
11401 ctgctgctca ccctgggtccg caccgaggcc gccgccgtcc tcggccacgc
11451 ctcggccgag tccttccccg agcgccgcgc cttccgtgac ctcggtctcg
15 11501 actcggtcac cgccgtggac ctgcgcaacc ggctcgtggc cggcaccgga
11551 ctgcggctgc cctcgacgat ggtcttcgac caccccaact gcgcggcgct
11601 cgccgcgttc ctgaagacga cggcgctcgg cgtccccggc gccgcaccgc
11651 agcagcacgc cgctaccggc accccggccg acgacgaccc gatcgccgtg
11701 atcggcata gctgccgcta ccccggcggc gccgccaccc ccgaggaaact
20 11751 gctgcggctc gccctcgacg gcgcgcacgt catctcggag tccccgcgg
11801 accgcccgtg ggacgcccgg ggctgtacg acccgaccc cgaccgcccc
11851 ggccacacct actcgtcca gggcggttc ctccacgagg ccgccggctt
11901 cgatcccggc ttcttcggga tctccccgcg cgaggcggtc gccatggacc
11951 cgcagcagcg gctcctgctg gagacctcct gggaggcggt cgaacgcgcc
25 12001 ggtatcgacc ccgcgtcact gcgcggcagc gccgccggca ccttcttcgg
12051 cgccagctac caggactact cctccaccgt gcagaacggc acgggggagt
12101 ccgaggcgca catggtgacc ggcaccgcgg ccagtgtcct gtccggccgg
12151 gtctcctacc tgctcggcct ggaggggccc gcggtcaccg tggacaccgc
12201 ctgctcctcc tcaactggctg ccctgcacct ggcctgccag tccctgcgcg
30 12251 acggcgagag ctccctcgcg ctggccggcg gtgcggccgt gatggccacc
12301 ccgcacgcgt tcgtcggctt cagccggcag cgtgccctgg ccaaggacgg
12351 ccgctgcaag ccgttctccg acaccgccga cggcatgacg ctcgccgagg
12401 gcgtcggcgt cgtcctgctg gagcgctgt cccacgccc cgccaacggg
12451 caccgggtac tggcgtgat ccgcggttcc gccgtcaacc aggacggcgc
35 12501 ctccaacggc ctgaccgcgc ccaacggccc gtcccagcag cgcgtcatcc
12551 gccaggcgct cgccaacgcc ggctgaccg gcgccgacgt cgacgcggtc
12601 gaggcgcacg gcaccggcac caagctgggc gaccccatcg agggccaggc

12651 cctgctcgcc acctacggcc aggaccgcga cgccgaacgg ccgctgctgc
 12701 tgggctcggg gaagtccaac atcggccaca cccaggccgc ggccggcgtc
 12751 gccggcgta tcaagatggt gctggccatg gacgccggcg aactgcccgg
 12801 caccctgcac ctcgacgcgc cctccagcca cgtcgactgg accgccggcg
 5 12851 ccgtggaact gctgcgcggg cgcaccccggt ggcccagagag cgggcgcccc
 12901 cgccggggcg gtgtctctc gttcgycatc agcggcacca acgcccacct
 12951 gatcctcgaa caggccccgg ccaccgagcc gccagccgac cccgaccgcc
 13001 tccgggacac cgccaccgac accgtcgtcc cctggccgct cgccgccaag
 13051 tccccggccg ccctgcgcgc ccaggccgcc cggctcctcg ccaccgtcga
 10 13101 gcacgacccc gacctccgc cgcaccccggt gggccacgcc ctggccacca
 13151 ccgcgcgcgc cctcgaaacac cgcgccgtcg tcgtcggcga gcgcgcgag
 13201 gacttctcgc gcggcctggc gcacctgtcc accggcgctt cgacggccgg
 13251 cctggtcagc ggcacgcgc gccccgaccc cgagggagcg gtcttctgtt
 13301 tccccggcca gggatcccag tgggagggaa tgggcccga actcctcgcc
 15 13351 acgtccgagg tgttcgcgc cgcgatcgat gactgcgcga cggccctcgc
 13401 ccggtacgtc gactggtcgc tgcacgacgt cctggccggc gagggcgacc
 13451 ccgcccgtct ggagcgggtg gacgtgggtc agcccgcgtt gttcgccatg
 13501 atggtcgggc tgtccgcgt ctggcgctcc cagggcgctg tccccggggc
 13551 cgtggtcggc cactcgcagg gcgagatcgc cgcggcctgc gtcgccggag
 20 13601 ccctcagcct ggccgacgcc gcccgcggtg tggcgctcgc cagccaggca
 13651 ctgccgaac tgtccggacg cggcggcgat atgtcgggtt cgcaccccggt
 13701 agagcgggtc accgcactcc tcgccccgtg gcaggaggcg ctgtccgtcg
 13751 ccgcgggtcaa cggccccctg tccgtgggtc tctccggcga caccgacgcg
 13801 ctcgacgccc tgcacaccgc ctgccaggaa cagggcgctg gggcccgcaa
 25 13851 ggtgtccgtg gactacgcct cgcacgggag gcacgtcgag gccgtccgcg
 13901 acgaactcgc ccgcgtcctc gcgcgggtcg acccgcgcg ccccgagggtg
 13951 ccgttctact cgacgggtcac cggcgaccgc gtggacgacg ccgccttcga
 14001 cggcgccctac tggtagacca acctccgcca gaccgtccgc atggaggagg
 14051 ccacccgcgc cctcctcgcc gccggacacc gcgtcttcat cgaggtcagc
 30 14101 ccgcacccgg tgcgcgcgc cccgatccag gagacgcagg aggcgtagc
 14151 ggagggcacc ggccgggtcc cgggtggctt cggctcgctc cgccgcgacg
 14201 agggcgggcc gcggcgcttc ctgacgtcgc tcgccgaggc ccacacccac
 14251 ggcgccccgg tcgactggac caccaccttc gcccggtccg cctaccagcc
 14301 ggtggacctg ccgacctacc ccttccaacg acaggacttc tggcccagg
 35 14351 cccggccgcg caccgccggc gccggcgccg acgcgtccga cgccgcgttc
 14401 tggcaactgg tcgagaacca ggacctcgcc gcgtcgccg acgcgtcgg
 14451 cgtccccgcc gacgacgagc acaccgcgtt cggcaccgtg ctgccggccc

14501 tgtccgcctg gcgcgccaag gccagggccc gcacccggat cgacgaactc
14551 cgctaccacg tccagtggac ccgggtcgcc gagcccgcg cgccccccac
14601 caccggccgg ctgctggctg ccgtcccgcc ggaccacgcc gacgccccct
14651 gggtcgccgc ggcgctcgac gccctgggca ccgacaccgt ccgcttcgag
5 14701 gccaaaggca ccgaccgcgc gggatgggccc gcacagatcg cccaactcgt
14751 cgaggacggc gaggagtcca ccggcggtgt gtcgctgctg gccgcccgcg
14801 aggatctcca cccggacttc ggctcggtac cgctggggct ggggcagacc
14851 ctcgtcctcg tccaggccct ccggcgacgc ggcctgaccg cgcctctgtg
14901 gtgcctgacc cgcggcgccg tcgccaccgg ccgcgacgac gccctcgaca
10 14951 gcccgaccca gggcgccctg tggggcctcg gccgggtcgt ggcctggaa
15001 caccgacgac gctggggcg cctgatcgac ctgcccgcc cctcgacgc
15051 ccgcgcgcg gccgcctca ccggcctgct cgcgacccc gccggtgagg
15101 accaaactcg cgtccgcgcc accggcggtc tcgcccgcg catggtgcac
15151 gccgcgcgt ccgcgccccg caccgggcgc cgctggcggg gccgcggcac
15 15201 ctgcctgatc accggcgcca ccggcgccat cggcgcccg gtcgcccgt
15251 ggatggccga gcacggcgcc gccacactgg tcctgaccag ccggcgcgcc
15301 ccggacgcac ccggcgccgc cgcactccgg gccgaactgg aggcctggg
15351 cgcccgctc accctcgccg cctgcgacgt cgcgaccgc gacgcctgg
15401 ccgcctgct gcccgacct cccgcgacc agccgctcac ctccgtcttc
20 15451 cactccgccg gcgtggccga cggggacgcc cgggcagccg acctgacct
15501 agatcagctc gacgcgtgc tcgcgcgcaa actgaccgcc gccaccacc
15551 tgcacgagct gaccgcccc ctcgacctg acgcgttcgt gctcttctcc
15601 tccggcgccg cggctctggg cagcggcgcc cagcccggct acgcgcccgc
15651 caacgcctac ctcgacgccc tcgcgcgcca ccgcaggtcc ctcgacctgc
25 15701 ccggcgcgct cgtcgccctg ggcacctggg gcgaggtcgg catggccacc
15751 gtccccgagg tccacgagcg actgcaccgc caaggggtcc gcgccatgga
15801 accggaccac gcgatcggcg cgctccagca gatgctggag gacgacgaca
15851 ccacctcgc cgtgacctc atggactggg aggcgttcgc accgagcttc
15901 accgcgaccc gaccagcgc cctgttcagc acggtgccc aagccgtccg
30 15951 cgcggtgacc ggcgacccgg gcaccacggc cggcgacgac gtggactccg
16001 cgaccccgcc gctccgcgc cacctggagg agctgtccgc cgcgagcgc
16051 ggccggggcc tggtcgaggc ggtccgcgcc gaggcgtccg cgacctcgg
16101 ccacgacacc cccgacgcca tcccgcgccc ccgtgccttc cgcgacgtcg
16151 gcttcgactc ggtcaccgcc gtcgaactgc gcaaccggct gcgcaccgcc
35 16201 ctcggcctgc cgctgcggc cgcgctcgtc ttcgaccacc cccccccac
16251 ggccctcgc ggccacctc gcgcgctgct ctcggcacc gcccccgagg
16301 acgcccggac cggccgcccc gacgacccc acgcccgcac ccgcgaggcg

16351 ctcgccaccg tccccatcgg acggtgcgc aaggcgggccc tcctcgacat
 16401 ggtgctgaaa ctcgccgacg gagacgcgac cgacgcccc gccccgagg
 16451 ccgacgcccc ctcggaatcc ctcgacgaca tggacgccga agccctgctg
 16501 cggctggcca ccgagaactc ggcgaactga aagagagctg gagccacca
 5 16551 tgagcacgaa ccccgacaag tacgtcgagg cactccggtc gtccctgaag
 16601 gagatcgagc ggctgcgccc gcagaacgaa cagctgggtg cggcgggcgt
 16651 cgagcccgtc gcggtcgtcg gcatcggtcg ccgcttcccc ggcggcgtca
 16701 cctccccga ggacctgtgg gagctggtcg ccgaggggcg cgacgtcatc
 16751 gggccgttcc cgcaggaccg cggtgggac ctggagaagt tggccggcgg
 10 16801 cggcgagggc ggcagcctcg cgcaggtcgg cggcttcgtc gaggacgccg
 16851 ccggcttcga ccccggttc ttcggcatct ccccgcgca agcggtcgcc
 16901 atggaccgac agcagcgcac cctgctggag atcacctggg aagccctgga
 16951 acgcgcggc atcgaccgt ccacctgac cggtaccccc accggcgtct
 17001 tcgtcggcac caccggccag gactacggcg aggtcatcaa ggcgtccgcc
 15 17051 gaggacgtcg aggtctactc gaccacggc cacgcgccca gcgtcatctc
 17101 cggccggctc tcctacacc tcggcgccga gggcccgccc gtcaccgtcg
 17151 acaccggctg ctccctcgcc ctggtcgccc tgcactgggc cgtccaggcg
 17201 ctgcgcggcg gcgagtgtc catggccctg gccggcgggc cgtccatcat
 17251 ggccaccccc ggcccgttcg tcgccttcac cgcgcagagc ggcctggccc
 20 17301 ccgacggccg ctgcaagccc ttctccgacc gggccgacgg caccggctgg
 17351 ggcgagggcg ccggcatgtt ggtcctgatg cggctctccg acgcccagcg
 17401 cgagggccgc ccggtcctcg ccgtgctgag cggctccgcc atcaaccagg
 17451 accggcgctc caacggcctg accgctccca accgccccctc ccagcagcgc
 17501 gtcacccgag ccgcgctgga cagcgcaccac ctacccgccc ccgacatcga
 25 17551 cgccgtcgag gcccacggca ccggcaccac cctcggcgac ccgacgagg
 17601 cccaggcgct cctggcgacc tacggacagg accggccgag gccctgtgg
 17651 ctcggctcgg tgaagtccaa catcgccac acccaggccg cctccggtgc
 17701 cgccggcgtg atcaaaatga tcatggcgtt gcagcgggac gtgctgcgc
 17751 gcagcctgca cgccaccgaa cccaccacgg acgtcgactg gaccgcccgc
 30 17801 tccgtcgacc tcctcgacga gacggtcgcc tggcccgaga ccggacgcgc
 17851 ccgcccgcgc ggcgtctcct ccttcggcat cagcggcacc aacgcccacg
 17901 tcacccctga acaggccccc accgcccccg aagagccac caccgaaccc
 17951 accgtccgac ccgcccgtgt cccgtgggag ctctccgccc gcaccgcgc
 18001 cgccctcgac gcccagcgcg ccgcctcac cggccacctc gccgacacc
 35 18051 ccgacgccga cccctcgac gtcggctacg cgctcgccga ccgacgcgc
 18101 accttcgaac accgcgccgt cctgtccccc gacggcaccg aactcgccca
 18151 cggaaccgcc ggcaaggcc cctgcgccgt cctcttctcc ggccagggt

	18201	cccagcgccc	gggcatggga	cgcgaactcc	acgcccgtt	cccgggtgttc
	18251	gccgcggcct	tcgacgagat	cacagcgctc	ctcgacaccc	acctcgaccg
	18301	cccgtcgcg	gaggctcgtct	ggggcaccga	cgccgacctg	ctgaacgaca
	18351	ccggctgggc	ccaacccgcc	ctgttcgccg	tcgaggtcgc	cctctaccgc
5	18401	ctggctcggt	ccctcggcgt	caccccgac	ttcgtcggcg	gccactccat
	18451	cggcgagctc	gccgcgcgc	acgtcgccgg	ggtcctctcc	ctcgaagacg
	18501	cctgcaccct	cgtcgccgcc	cgcccccgc	tcatgcaggc	cctgccgcgc
	18551	ggcggcgcg	tgctcgcgat	ccgcgccacc	gaggacgagg	tcacccccca
	18601	cctcaccgac	gacgtctcga	tcgccgccgt	caacggggcc	acctccgtcg
10	18651	tcgtcgccgg	caccgaggaa	gccgtcgccg	cgatcggggc	gcgcttcacc
	18701	gccaggacc	gcaagaccac	ccggctcgcg	gtcagccacg	ccttcactc
	18751	gccgtcatg	gacccgatgc	tggcggaatt	ccgcgcgcgc	gccgcgggcc
	18801	tgacctacca	cgagccgcgc	atcccggtcc	tctccaacct	caccggcacc
	18851	gtcgccgccg	tcgccgacct	gtgctccgcc	gactactggg	tcgccacgt
15	18901	ccgcgaggcg	gtccgcttcg	ccgacggcgt	caccgccctc	accgaccgcg
	18951	gcgtgaccac	gctcgtcgaa	ctcgggcccg	acggcgtgct	gtccgccatg
	19001	gcccgaggaat	ccctgccgga	cggcgccgcc	gccgtgccgc	tgctgcgcaa
	19051	ggaccgcccc	gaggagctct	ccgcgtcac	cgccctggcc	cgcccccacg
	19101	tccgcggcgt	cacggctccg	tgggccggcc	tcttcgacgg	caccggcgcg
20	19151	cgccgcgccg	acctgcccac	ctacccttc	cagcaccagc	ggttctggcc
	19201	gaccgcggcc	cgcgcgcgcc	aggacgtcac	cgccgcggga	ctgggcgccg
	19251	ccgaccaccc	gctgctcggc	gccaccgtcg	aactcgccga	cggggccggc
	19301	tacttgttca	ccagccggct	ctccgtccgg	accacccct	ggctcgccga
	19351	ccacggggtc	cagggccggg	ccctgctgcc	cgccaccgcc	ttcgtcgaac
25	19401	tggccgtccg	cgccggcgac	gaggccggct	gcgaccgcgt	cgaggaactg
	19451	accctggccg	ccccctggt	gctgcccag	cgcgggcgcg	tccaactcca
	19501	ggtccgcgtc	ggcgcgcgcg	acgcgcgcg	ccgcgcacc	ctcggcatct
	19551	tctcccgct	cgaggacggc	ttcgacctgc	cctggtcgca	acacgccacc
	19601	ggcgtcctga	ccgcggcg	cgccgcgc	gacccacct	tcgacgccac
30	19651	cgtctggccc	cccagcgggc	ccgaaccgt	cgacctcacc	ggcgcgtaag
	19701	agcgccctgg	cgcactcggc	ttccagtacg	gccccgcctt	ccagggcctg
	19751	cgcgcgcgct	ggcgccgcga	caccgaggtc	tacgccgaag	tggccctgcc
	19801	cgacggcgcg	gacaccgacc	ccgcgcctt	cggactgcac	ccggccctgc
	19851	tggacgcgc	acaacacgcc	gcccctacg	ccgacctcgg	cgccatcagc
35	19901	cgcggcgggc	tgccgttcgc	ctgggaaggc	gtctcgctcg	ccgcgcgcgg
	19951	cgccaccacc	gtccgcgcgc	ggatcgcccc	ggccggcgag	gacaccgtca
	20001	ccatcgccgt	ctaagacgcc	gccggcgcca	ccgtgctgtc	cgtcgactcc

20051 ctggtctccc gcgaggtccc egccgacgca cccggcgccg ccggcacagt
20101 ccaccgagac tccctcttcc acgtcgagtg gaccccgctc cagggcgccc
20151 cgggccccgc accggccacc gtcgccgtcc tcggccccga cccggacgcc
20201 ctcgccgaca ccctccgcgc caccggcatc cggaccaccg cccccgcga
5 20251 cctggccgcc ctcgccgacg ccgaagggcc cgtccccgac ctggtcgtca
20301 ccaccctcac caccaccccg ggcgcccccg tccccgacgc cgcgcacgcc
20351 accaccgccg ccgtcctcgc cctcgcccaa cagtggctcg ccgacgaccg
20401 cttcgccgac gcccgcttgg tctcgtcac ccgcggcgcc accgacggca
20451 ccgacccccg cgcgcgggcc gccggcggcc tgatccgcac cgcctgcacc
10 20501 gagaaccccc gccgtttcgc cctcctcgac ctcgcccccg acaccggccg
20551 gcccgacccc gagaccctgg ccaccgccct ggccgccagc caccgacgac
20601 ccgacctcgc cgtccgcggc accgacgtgc acgcgcgccg cctggcccgt
20651 gtcccgtcgc ccaccgaacc caccacctgg aaccggacg gcacctcct
20701 gatcaccggc ggcaccggcg gcctggggcg ggctcctcgc cgcacctgg
15 20751 tcgccaccca cggcgtccgc cacctgctgc tcgccagccg ccgcggccccg
20801 gccgccgacg gcgccgacga cctgacggcc gaactcaccg ggctcggcg
20851 caccgtccac atcgccgcct gcgacgtcgc cgacccggcc gccctcgcg
20901 acctgctcgg caccgtcccc gccgggcacc cgtcaccgt cgtcgtccac
20951 accgccggcg tcgtcgacga cggcgtcctc ggctccctca cccgcagcg
20 21001 cctggacacc gtctcgccg ccaaggccga cgcgcctgg cacctgcacg
21051 aggcgacccc ccacctcgac ctggacgcct tcgtcctctt ctcgctccgc
21101 gccgccaccc tcggcagccc cggacaggcc aactacgccg ccggcaacgc
21151 ctctctggac gccctcgcg cccggcgcg cgcaccggc ctgcccgcga
21201 cctccctcgc ctggggcccc tggaccaga gcgtcggcat gacaagcagc
25 21251 ctgtccgacc tcgacgtcga gcgcacgcc cgtccggca tgccccgct
21301 gaccctggaa cagggcaccg ccctcttctga cgcggccctg gccgccggcg
21351 ccgccgccct cgcgccggc cgcctcgacc tgcccgtcct gcgcaccag
21401 ggcgacatcg ccccgctgct gcgcggcctg atccgcaccc ccgtgcggcg
21451 caccgccgcc caggtctcgc agaccgccga cggcctcgc cagcggctcg
30 21501 ccggcctcga cgcgcgcgc cgcgggaag ccctcctgga actcgtccgc
21551 acccagatcg cccaggctct cggccacgcg gacgccaccg aggtggagac
21601 cggccgccag ttccaggacc tcggcttcga ctccctcacc gccgtcgaac
21651 tccgcaacgc cctgaacacc gccaccggcc tgcggctgcc cgcaccatg
21701 gtgttcgact acccgacacc acacgccctc gccgaccacc tgcgcgacga
35 21751 actcctgggc accgaggccg agtcgaccac cgcctcccc gtgccgaccc
21801 gtaccgccgg caccgacgac ccgacgtca tcgtcggcat ggctgcgcg
21851 taccgccggc gcacgcctc acccgaggac ctctggcgcc tggtcagcca

21901 gggcgccgac gccactggcc cgttccccac caaccgcggc tgggacctgg
21951 acaacctcta cgaccccgac cccgaccgcc cgggcccgcac ccacgtccgc
22001 gccggcggct tcctgcacga cgccgggtcc ttcgacgccg acttcttcgg
22051 gatgagcccc cgcgaggcga tggccaccga ctcccagcag cgcttctgctc
5 22101 tcgaactctc ctgggaagcc gtcgaacgcg ccggcatcga ccccgccctca
22151 ctgcgcgact ccggcaccgg cgtcttcgcc ggcgtcatgt acaacgacta
22201 cggcaccacc ctgaccggcg acgagtacga ggcgttccgc ggcaacggca
22251 gcgccccgag cgtcgcctcc ggccgcgtct cctacaccct cggcctggaa
22301 ggccccggcg tcacggtgga caccgcctgc tcttctcttc tggtcgcctt
10 22351 gcaactggcg gcgcaggcgt tgcgggcggg ggagtgtctg ttggcgttgg
22401 ccggtggtgt gacggtgatg tcgacgccga gcacgttcgt ggagtctctg
22451 cggcagcggg gtctggcgcc tgatggctgt tcgaaggcgt tcgccgaggc
22501 cgcggacggc gtggcctggt ccgagggcgt cggcatgctg gtcctggagc
22551 ggcagtcgga cgcggtgcgc aacggtcacg agatcctggc cgtggtgcgc
15 22601 ggctcggcgg tcaaccagga cgggtgcgtcc aacggtctga ccgcgcccac
22651 cggcccgctc cagcagcggg tgatccgtca ggcgttggcc agtggcggcc
22701 tgtccacggc cgacgtggac gccgttgagg cgcacggcac gggtagcagc
22751 ctcggtgacc cgatcgaggc ccaggcgtc ctggccacct acggtcgcga
22801 ccgcgacccc gagaaccgc tgctgctcgg ctcgatcaag tccaacatcg
20 22851 gtcacaccca ggcagcggcc ggtgtcgccg gtgtcatcaa gatggtcatg
22901 gcgatgcggc acggcgtgct gccgcagacc ctgcatgtcg acgcgccgtc
22951 ctgcacgtc gattggagcg tcggcgcgt cgaactgtc accgagcaga
23001 ccgcctggcc ggagaccggc cgggcccgtc gcgccggtgt ctctctcttc
23051 ggcattcagc gcaccaacgc ccacgtcgtc atcgagcagt ccccgaccgc
25 23101 cgtccccgcc acgcccgcgt ccgccgaccg gtccgtcgag gaaccgcccg
23151 ccgtcccctg ggccctgtcc ggcaagacct ccgacgccct ccgcgaccag
23201 gccgcccgcc tcctcgccca cgtcgaggcc caccgccac tgcgcccgt
23251 cgacatcagc tactccctga tcgccaccgc caccgccttc gaccaccgcg
23301 ccgtcgtcct cggcaccgac cgcgccgagg ccctgcgcgc cctcaccgcc
30 23351 ctgcgcgccg gcgagaccga cccggccgcc ctaccggca ccgtccgcac
23401 cggccgcacc gccttctct tctccggcca gggctccaa cggctcggca
23451 tgggcccgt cctctacgag cggttccccg ccttcgccga agccctcgac
23501 accgtcctca ccgccctcga cgcggaactc ggccaccccc tccgcgacat
23551 catctggggc gaggacgtc aactcgtcga ccggaccggc tacaccaaac
35 23601 ccgcctgtt cgccatcgag gtggcactct tccgcctcct ggaagcctgg
23651 ggcacacac cggacttcgt ggccggccac tccatcggcg agatcgccgc
23701 cgcacacgtc gccggcgtgc tctccctcgg cgacgcctgc cgcctcgtcg

	23751	tgccccgcgc	cgtgctgatg	cagtcgctgc	ccgaaggcgg	cgcgatgac
	23801	gccgtccagg	ccaccgagga	cgaggctctg	ccccctctca	ccgacgacgt
	23851	ctcgatcgcc	gccgtcaaca	gcccgcacct	cgtcgctcgtc	tccggctacg
	23901	agaacgccac	cctcgccgtc	gcccggcact	tcgccgacca	gggcgcgcgc
5	23951	accacgcggc	tgcgcgctcag	ccacgccttc	cactcgccgc	tgatggcgcc
	24001	gatgctcgac	gacttccgcg	ccgtcgctcga	gagcctcacc	ttcaccgccc
	24051	ccacgacccc	cgtcgtctcc	aacctgaccg	gcgaactggc	cccggccgag
	24101	gcgctctgct	cggccgacta	ctgggtccgg	cacgtccgcg	aggcggctccg
	24151	cttcgcccgc	ggcatccgca	ccctcgccga	ccgcggcgctc	accaccttcg
10	24201	tcgaactcgg	ccccgacagc	gtgctgtccg	ccatggccca	ggagtccgcc
	24251	cccgaaggcg	ccggcaccat	cccgtcctg	cgccgcgacc	ggcccgcagga
	24301	acaggccgtc	ctggccgccc	tctgccacct	ccagggtgctc	ggcgctcgagg
	24351	ccgactggtc	cgccaccttc	cgcggcctcg	accccgctccg	cgctcgacctg
	24401	ccgacctacg	ccttccagca	ccgtggttc	tgccccgccc	cccgcaccgc
15	24451	ccgccccgac	gacgtccgcg	ccgcgggcct	gggcgcgcgc	gaacaccccc
	24501	tcctcggcgc	cgccgtgcaa	ctccccgacg	acgacggcgc	actcttcacc
	24551	ggccgcctct	ccctgcgcac	ccaccgctgg	ctggccgacc	acaccgtcct
	24601	gggcaccgtc	ctgctcccgg	gcaccgcact	ggtggaactc	gccgtccgcg
	24651	cgggcgacga	gaccggcagc	ggccacctcg	aagaactcac	cctcgccgcg
20	24701	cccctgaccc	tccccgagga	cggcgccacc	ctcctccagg	tccgcgtcgg
	24751	atccgccgac	gacaccggcc	gccgcaccgt	caccgtccac	ggccgccccg
	24801	acgacaccgc	cgaccgcacc	tggaacgctgc	acgccaccgg	tgtgctcgcc
	24851	accacgccac	cggccgcgcg	ggcgcttcgac	accacggtct	ggccgccccg
	24901	cgacgcgaa	cccccaacca	ccgacgactg	ctacgcacac	ttcaccaccc
25	24951	accgcttcgc	ctacggcccc	gccttccagg	gcctgcgggc	cgcttgccgc
	25001	gccggcgacg	tgctgtacgc	cgaggctgcc	ctgccggagt	ccgccaccga
	25051	cgaagcggcc	gccttcggcc	tgcacccggc	gtcctggac	gccggcctgc
	25101	acgccgcgct	cctcgccgac	gaccgcgaca	ccggactccc	gttctcctgg
	25151	gaaggcgctca	ctctgcacgc	ctccggcgcc	accgcgctac	gcgtccggct
30	25201	cgccccgaac	ggccccaaac	gcctgtccgt	caccgccgcc	gaccgggccg
	25251	gcaaccccg	cgccaccgtc	acccgcctgc	tcgccgccc	cctggacgcc
	25301	gagcagttga	ccatccacag	cgccctgacc	cgcgacgcgc	tcttcacct
	25351	ggactggacc	ccggtcccgc	ttcccgacac	cgccaactcc	gcgcgcgcgg
	25401	ccctcctcgg	cccggacacc	gccgtgctcg	ccgacgccct	cggcgacccg
35	25451	gccgtcgcac	gccacgcaac	cctcgacgac	ctcctggccg	gggacaccac
	25501	cccgcgcgc	acggtcctcg	tccccctcgg	cgccccactc	gacggcgaca
	25551	ccgcgcagca	cgcgcacgcc	ctcaccgcga	gcgcgctgac	cctcgctccag

25601 cagtggctcg ccaccgaccg cctcgccgac tcccgctgg tcttcgtcac
25651 ccaeggagcc gtcgccaccg acgacgcgcc cccaccgac ctggccgccg
25701 ccgcggtctg gggcctgatc cgctccgcgc agaccgagaa ccccggcacc
25751 ttcacctcc tcgacctga caccgagccc gactcgacca ccgcgctcag
5 25801 ccgcgccctg accctcgacg aaccacagct cctcctccgc gccggccgcg
25851 cccgcgcgcg ccgcctcacc cgcacccccg cccccaccac caccaccac
25901 acgccgtggt ccgcggacgg aacggtggtg gtgacgggtg gtacgggtgg
25951 tctgggtggg ttggtggccc ggcattctgt gcggtcgtgt ggggtgcggc
26001 atttgttgtt gaccagtcgt tctggtgtgg gtgctgcggg tgcggccggg
10 26051 ttggtcgcg agttggagtc gttgggcgcg cgggttgtgg ttgcggcgtg
26101 tgatgtgggt gatggctcgg ctgttgcgga gttggttccc ggtgtgtcgg
26151 agtcgtatcc gttgtctgcg gtggtgcatg cggctggtgt gttggatgac
26201 ggtgtggtgg gttcgttgac gccggagcgg ttggctgcgg tgttcgtcc
26251 gaagtggtat ggtgcgtgga acctgcatga ggcgacgcgt ggtctggatc
15 26301 tggacgcgtt tgtgtcttc tcgtctgttg cgggtgtgtt cgggggtgcg
26351 ggtcaggcca actatgcggc gggtaatgcg tttttggacg cgttgatggt
26401 tcacggggtg gctggtgggt tgcctggtgt gtcgttggcg tggggtgctt
26451 gggatcaggg tgtggggatg acggcggggc tgacggagcg ggatgtccgt
26501 cgtgctgctg agtcgggtat gccgttgttg acggttgatc aggggtgtggc
20 26551 gttgttcgat gcggcggttg cgacggggag tgccgcgttg gtgccggtcc
26601 gtctggacct ggccgcaact cgcacccggg gcgacatcgc accgctcttc
26651 cgcggcctcg tccgcgcacc gctgcgcgc accgcggcca ccggcctcgc
26701 caccggcgcg gacaccggcc tcgtccaacg gctcggccga ctcgaccacg
26751 cccaacgcca cgaggcaact ctcgacatgg tccgcagcag cgcgcgctc
25 26801 gtcctcggcc acgccgacgg caacgccatc gacgccgaac gcgccttccg
26851 cgacctcggc ttcgactcgc tcaccgcggt cgaactccgc aaccgtctgc
26901 gcaccgccac cggcctgcac ctgtcggcca ccatggtctt cgaccacccc
26951 accctgtccg ccctcgcgga gcacctgcgg gacgagttgt tcggcgcggt
27001 cgagagcgag gtgcgggtgc cgtccaggc actgcgcgg accgcggacg
30 27051 atcccatcgt ggtggtgggc atggcctgcc gtttccccgg tgggtgtgacc
27101 tcgcccaggg acctgtggcg cctggtcgac gacggcaccg acgccatcac
27151 caccttcccg accaaccgcg gctgggacct ggacaacctc tacgaccggg
27201 accccgagca cttcggcacg tcgtacaccc gtcgggtgg cttcctgcac
27251 gaggcggggg agttcgaccc ggcgttcttc ggaatgagcc cgcgtgaggg
35 27301 gctggcaacc gactcccaac agcgtctcct gctggaatcc tcctgggagg
27351 cgatcgagcg ggccggcatc gaccgcgtga ccctgcgcgg cagcgccacc
27401 ggcgtcttcg ccggcggtgat gtacagcgac tacgggagca tcctcggcgg

27451 caaggagttc gagggcttcc aaggccaggg aagtgcgggc agcgtggcct
 27501 cgggccgcgt ctccctacgc ctcggcttcg agggcccggc cgtcacggtg
 27551 gacacggctt gctcttcctc tctggtcgcc ctgcactggg cggcgcaggc
 27601 gttgcgggcg ggggagtgct cgttggcgtt ggccgggtgt gtgacgggtga
 5 27651 tgtcgacgcc gagcacgttc gtggagtctt cgcggcagcg gggctctggcg
 27701 cctgatggtc gttccaaggc gttcgccgag gccgcggacg gcgtcggctg
 27751 gtccgagggc gtcggcatcc tcgtcctgga gcgccagtcg gacgcgggtgc
 27801 gcaacggcca cgagatcctc gccgtgatcc gcggctcggc ggtcaaccag
 27851 gacggtgcgt ccaacggcct gaccgcgccc aacggcccgt cccagcagcg
 10 27901 cgtcatccgt caggcgttgg ccagtggcgg cctgtccacg gccgacgtgg
 27951 acgccgtcga ggcgcacggc acgggtacga cgctcgggtga cccgatcgag
 28001 gcccaggcgc tcttgccac ctacggccgt gaccgcgacc ccgagaaccc
 28051 cctgtggctg ggctccctga agtccaacat cgggcacacc caggcagcgg
 28101 ccggtgtcgc cgggtgcctc aagatgggtc tggcgatgcg gcacggcgtg
 15 28151 ctgccgcaga ccctgcatgt cgacgcgccg tctcgcacg tcgattggag
 28201 cgtcggcgcc gtcgaactgc tcaccgagca gaccgcctgg ccggagaccg
 28251 gccgggtccg tcgcgcgggt gtctcctcct tcggcatcag cggcaccaac
 28301 gccacgtca tcgtggaaca gccggcgctc gtcgaaagcc cggccgcgga
 28351 gccgagcggg cgcgaacccg gcgtcgttcc gctgccgctg tccggaaagt
 20 28401 cccccgaggc cctgcgcgac caggccgcac gcctgctggc cgggttggcg
 28451 gagcggcccc cgctgcgccc gctcgacctc ggctactcgc tggcgacgac
 28501 ccgttcggcg ttcgaccacc gggcgggtgt gctcgccacc gaccgcgccg
 28551 atgcggtccg cgcgctgacg gcgctcgccg ccgccgacgc ggatctctcc
 28601 gccgtcgtcg gcgacacccg caagggtcgt cagcgggtgc tgttctcggg
 25 28651 tcagggtcgc caacgcctgg gcatggggcg tgagttgtac gagcgtttcc
 28701 cggctctcgc cgaggctctc gatgtcgcga tcgaccacct ggacgccgcc
 28751 ttgccgccc aggccagtct cgtgaggtg atgtggggcg acgatgtcga
 28801 gctgctggac gagacgggtt ggacgcagcc ggctctgttc gccgtcgagg
 28851 tggccctgtt ccggtcgggt gagagtggg gtgtccgtcc ggacttcgtg
 30 28901 gccggtcatt ccacggtga gatcgggcg gcgcatgtcg tcggggtgtt
 28951 ctcgctggag gacgcctgcc gtctgggtgg cggccgtgcg acgctgatgc
 29001 aggcgctgcc gaccggcggc gcgatgatcg cgatccaggc cgccgaggac
 29051 gaagtcaccc agcacctgac tgacgacgtc tcgatcgccg ccgtcaacgg
 29101 cccgacctcc gtggtcgtct ccggggccga gagcgtgcc cgcacgggtg
 35 29151 ccgaccggct cgcggagaac ggccgcaaga cgaccggct gcgggtctcg
 29201 catgcgttcc actcgccgtt gatggatccg atgctggcgg agttccgtgc
 29251 ggtggccgag ggctgtcct acgccacccc gaccctcccc gtcgtctcga

29301 acctgacggg ccggctggcc acggccgatg acctctgctc ggccgagtac
29351 tgggcgcgcc acgtccgcga ggccgtccgc ttcgccgacg gcgtcagcac
29401 cctggagaac gagggcgta ccacgttcct ggaactggga ccggacggcg
29451 tgctgtccgc catggcccag cagtcgctca ccggcgacgc cgccaccgtc
5 29501 ccggccctcc gcaaggaccg cgacgaggag acgtccgcgc tcaccgcctt
29551 cgcccacctc cacacggcag gtctccgcgt cgactgggcg gcgttcttcg
29601 ccggcagcgg cgccaccgcg gtggacctgc cgacctacgc cttccagcac
29651 gccacctact ggcccaccgg caccctgccc accgcgcacg ccgcggccgt
29701 cggcctcacc gccgcggagc acccgctgct gaacggttcg gtogaactcg
10 29751 ccgaaggcga aggggtgttg ttcaccggac ggctgtcact gcagtacat
29801 ccgtggctgg ccgaccacgc cgtcatggga caggctctgc tgcccggcac
29851 cgactgctg gaactggcgt tccgggccgg cgacgaggcc ggttgcgacc
29901 gcgtcgagga actgacgctc gccgcaccgc tcgtcctgcc cgagcgcggt
29951 gcggtacaga ccaggtccg ggtcggcgtc gccgacgaca ccggccgcgg
15 30001 taccgtcacc gtccactccc ggcccagaca cgcgaccgac gtgtcgtgga
30051 ccagcacgc gaccggcacc ctgaccatgg gctccgcccc ggccgacacc
30101 ggtttcgacg cactgcctg gccgcccgcg gacgccaac ccctcgccac
30151 cgacgactgc tacgcgcgt tcacgacgt cggcttcgcc tacggccggg
30201 tcttccaggg cctgcgggcc gcctggcgcg ccggtgacgt gctgtacgcc
20 30251 gaggtggccc tggcgagtc caccggcgac gaggcgaccg ccttcggtct
30301 gcaccccgca ctgctcgacg ccgccctgca cgcctccctc gtcgcccacg
30351 agggcgagga gagcaacggc ggactgccgt tctcctggga gggcgcgacc
30401 ctctacgca ccggcgccac cgcgctgcgc gtccggctga ccccgacggg
30451 caccgacggc cgttcggtgg ccatcgccgt ggccgacacc gccggtcgtc
25 30501 cggtcgccc catcgacaac ctgctctgc gccgggtctc cggcgaccag
30551 ttgaccggcg ccgcgggact ggcccgcgac gccctgttca ccctggactg
30601 gaacccgta ccggagaacc tcgtaccgga gaacctgta ccggagaaca
30651 ccggcggggg ccacgcccag gaccaggacg gccggcccgc cgcgccacc
30701 gtcgcgctgg tcggcgcgga cggcaccgcg atcgccgccc acctgaccgc
30 30751 cgccggcatc cacaccacc tccacccga cctcaccacc ctgcccagca
30801 ccgacgccga cgttccgaag acggtectca tccccctcac cggaaccgga
30851 accggaaccg gcaccgggac tgagtcgacg gacggaatcg ggacgggggc
30901 cgccgagtcg gacgcgtccg cccctcccc ggccgaggtc gcccacccc
30951 tgtccaccgc cgactcgcc ctgctccagg agtggaccgc acaggagcgc
35 31001 ttcgcccgt cccgcctggc gttcgtcagc accggggcga cgcccgccgg
31051 cggtagcgac gtcattggac tggccgcccgc cgcggtctgg ggctgggtcc
31101 gatccgcca gtccgaagcc ccggacacct tcgtcctgat cgaccgtgac

	31151	cccggccccg	ccggcacgca	cgaccgcaca	gccgccgccg	aacggggcca
	31201	actgctccta	cgggcactgc	acaccgacga	accgcagctc	gccctgcgtg
	31251	acggcggcgt	gctcgccgcc	cgcttgcccc	gcttcgacac	cgcgcccgcg
	31301	ctcaccgccg	cgcccgaccg	ggcctggcgg	ctcgacagca	cggccaaagg
5	31351	cagcctcaac	ggcctcgccc	tgaccccgta	tccggcgcca	ctggcgccgc
	31401	tcaccggcca	cgaggtgcgg	gtcgaggtgc	gtgccgcggg	cctgaacttc
	31451	cgtgacgtgc	tcaacgcggt	ggggatgtat	ccgggtgatg	atgtcggatc
	31501	gttcggttcg	gaggcgcccg	gtgtggtcgt	cgaggtcgga	ccggaggtga
	31551	ccggcctggc	ccccggcgac	caggtcatgg	gcatgatcac	cggcagcttc
10	31601	ggctcgctcg	ccgtggacga	cgcgcgccgc	ctcgcccgcc	tgcccagga
	31651	ctggctcctg	gagacgggtg	cgtcggtgcc	gttgggtgtc	ctcaccgcgt
	31701	actacgccct	gaaggagtgt	ggtggtctgc	gggcggggga	gaagggtctg
	31751	gtgcatgccg	gtgcccgttg	tgctgggtatg	gcggcgatcc	agatcgcccc
	31801	gcatgtcggg	gccgaggtgt	tcgccacggc	cagtgaaggc	aagtgggacg
15	31851	tgttgcgctc	gctcggcgtg	gccgacgacc	acatcgccctc	ctccccgacc
	31901	ctcgacttcg	aggcggcctt	cgccgaagtc	gccggcgacc	gcggcctgga
	31951	cgctgccttc	aactccctcg	ccggtgactt	cgtcgacgcc	tcgatgcggg
	32001	tgctcgccga	cgccggcccg	ttcctggaga	tgggcaagac	cgacatccgc
	32051	gccgcggact	ccgttcccg	cgccctctcc	taccagtcct	tcgacctcgc
20	32101	ctgggtgggtg	ccggaaacca	tcggcaccat	gctggccgag	ctgatggacc
	32151	tcttcgcac	cgccgcactg	cgccactcc	cggtccgcac	ctgggacgtc
	32201	cggcacgcca	aggacgcgtt	ccgcttcattg	agcatggcca	agcacatcgg
	32251	caagatcgtg	ctcaccctgc	cccgtcctcg	gaagcccag	ggaacggtgt
	32301	tggtgacggg	tggtacgggt	ggtctgggtg	ggttggtggc	ccggcatctg
25	32351	gtgcggtcgt	gtggggtgcg	gcatttgttg	ttgaccagtc	gttctggtgt
	32401	gggtgctgcg	ggtgcggccg	ggttggtcgc	ggagttggag	tcgttgggcg
	32451	cgcggttgtg	ggttgccggc	tgtgatgtgg	gtgatggctc	ggctgttgcg
	32501	gagttgggtg	ccggtgtgtc	ggagtcgtat	ccgttgtctg	cggtggtgca
	32551	tgccgctggg	gtgttggtg	acggtgtggt	gggttcgttg	acgccggagc
30	32601	ggttggctgc	ggtgttgctg	ccgaagggtg	atggtgcgtg	gaacctgcat
	32651	gaggcgacgc	gtggtctgga	tctggacgcg	tttgttgtct	tctcgtctgt
	32701	tgccgggtgtg	ttcgggggtg	cggttcaggc	caactatgcg	gcgggtaatg
	32751	cgtttttgga	cgcggtgatg	gttcacccgg	tggctgggtg	gttgccctgg
	32801	gtgtcggttg	cgtggggtgc	ttgggatcag	gggtgtggga	tgacggcggg
35	32851	gctgacggag	cgggatgtcc	gtcgtgctgc	tgagtcgggt	atgccgttgt
	32901	tgacgggtga	tcagggtgtg	cggttggtcg	atgcggcggt	ggcgacgggg
	32951	agtgccgcgt	tggtgccggg	ccgtctggac	ctggccgcac	tgccgacccg

33001 gggcgacatc gcaccgctcc tccgcggcct ggtcaaggcg cccatccgcc
33051 gcgcagccgc caccacaccc ggcgacaccg gactcgccga gcagctcacc
33101 cggctccagc gcgccgagcg acgggacacc ctccctcgcg tcgtccgcga
33151 ccaggccgcg atggtcctcg gccacacctc gggcgacggc gtcgacccgt
5 33201 cccgcgcctt ccgcgacctc ggcttcgact cgctcaccgc ggtcgaactc
33251 cgcaaccgca tcggcgcggc caccggcctg cggctaccgg ccacggccgt
33301 cttcgactac cccaccgccc atgccctcgc cgcacacctg ctccaccgaac
33351 tgctcgcccc cgacgccgag tcggaccccc acgagccccg cgacccccacc
33401 gcgggaccga ccgacgaccc catcgtcac atcggcatga gctgccgctt
10 33451 ccccggcgac atcggctcgc cggaggacct gtggcgctg ctcggcgacg
33501 gcgccgacgt cgtcaccgac tccccgacca accgcggctg ggacctggac
33551 aacctctacg accccgaccc cgcgcacgcc ggcaacctcg acgcccgcac
33601 cggcggtttc ctgcacgacg ccgccgactt cgacgccgac ttcttcggca
33651 tgagcccccg cgaggccatg gccacggact cccagcagcg cctgctgttg
15 33701 gagtcctcgt gggaggcgat cgagcgggcc ggcatcgacc cgctgacctt
33751 gcgcgacagc cgcaccggcg tcttcgccgg cgtcatgtac agcggctacg
33801 gcacccgcct cgacggcgcc gaattcgaag gcttcaggcg gcagggcagc
33851 gcaactgagcg tggcctccgg ccgggtctcc tacaccttcg gcttcgaagg
33901 cccggccatg acggtcgaca ccgcctgctc ctccctcgctg gtcgccctgc
20 33951 acctcgccgc acaggcactc cgcggcggtg agtgaccctt cgccctcgcc
34001 ggtggtgtca ccgtgatgtc catcccgac accttcacg agttctcccc
34051 gcagcgcgga ctggcccccg acggccgctc caagccgttc tccgagtcgg
34101 ccgacggcgt cggctggtcc gagggcgctg gaatgctgct cctggagcgc
34151 cagtcggacg ccgtgcgcaa cggccaccag atcctggccg tggcgcgcg
25 34201 ctcggcggtc aaccaggacg gtgcgtccaa cggcctgacc gcgcccacg
34251 gcccgctcca gcagcgggtg atccgtcagg cgttggccag cggcggcctg
34301 tccacggccg acgtggacgc cgtcgaggcg cacggcacgg gcaccacgct
34351 cggtgacccg atcgaggccc aggccctcct ggccacctac ggccgcgacc
34401 gcgacccccga gaaccgctg ctgctcggtt cgatcaagtc caacctcggc
30 34451 cacaccaggg cagccgccgg tgctcgccgg gtcataaga tggatcatgg
34501 gatgcggcac ggcgtgctgc cccgcagcct gaacatcacc gagccgtcct
34551 cgcacgtcga ttggagcgcc ggccgctcg aactgctcac cgagcagacc
34601 gcctggccgg agaccggccg ggcccgtcgc gccggtatct cctccttcgg
34651 catcagcggc accaacgccc acgtcactct ggagcagccg gaggccgcgc
35 34701 ggcactcggc gccggaagaa gccgacacgg cggaggcagc cgccaaggcg
34751 ccggccaccg cgcacctgcc cgtaatgcc tgggcaactgt ccggcaagac
34801 gccggaggcc ctgcgtgccc aggccgcacg cctcctcgcc cacctccagc

34851 agcgccccga actcgacacc gccgacatcg ccctgtccct cgccaccag
 34901 cgctcccagt tcaccaccg ggcagtcgtc ctgagcaccg accgtgacga
 34951 ggcgaccgc gcgtgtccg ccctcgccac caccgcccgc tccgaccct
 35001 cggccctcac cggcacggc accatgggac gttgcgcggg gctgttctcg
 5 35051 ggtcagggct cgcaacgtct gggcatggg cgtgagttgt acgagcgttt
 35101 cccggtcttc gccgaggctc tcgatgtcgt gatcgatcac ctggacgccg
 35151 ccttgcccgc ccaggccggg ttgcgtgagg tgatgtggg cgacgatgtc
 35201 gagttgctga acgagacggg ttggaccag cccgcgctct tcgccatcga
 35251 ggtggcgctg tttcggttg tggagagttg ggggtgtccg ccggacttcg
 10 35301 tggccggtca ttccatcggg gagatcgcg cggcgcatgt cgtcgggggtg
 35351 ttctcgttg aggacgcgtg ccgtctggg gccgcgcgg cgacgctgat
 35401 gcaggcggtt cgggccggg gcgcgatgat cgcgggtccg gcgaccgagg
 35451 acgaagtcat ccgcacctg accgacgagg tggcgatcgc ggccgtcaac
 35501 ggcccacact ccgtggtgat ctccggcgca gaagaggcca cgcagaccgt
 15 35551 ggcacaacac ttccgccacc aggggcgcgc gacgaccgc ctgcgggtct
 35601 cgcgtgcgtt ccactcgccg ctgatgatgc tggcggagtt ccgtgcggg
 35651 gccgagggcc tgctctacgc caccgcacc ctcccgctc tctcgaatct
 35701 gacgggcccag gtggccacgg ccgacgaact ctgctcgcc gagtactggg
 35751 tgcgccacgt ccgtgaggcg gtccgcttcg ccgacggtgt gacggccctc
 20 35801 gaagccgagg gcgtgcggac cttcctggaa ctccggcccg acggcgtcct
 35851 cgccgccatg gccaggaaa ccgtcgccga cgacacggc accgtccccg
 35901 tcctccgcag gaacatgcc gaggaacgga ccctgctcac cgcactcggc
 35951 cggtccaca ccaccggaac ccgatcgac tgggcccgc tcctggcccc
 36001 gaccggcgcc cgcccggtg acctgccgac atacgcgttc caacaccgtc
 25 36051 ccttctggcc ctccggcccc cgcgacaccg cggatgccgc cgccgtcggc
 36101 atcgccggcg cgagccacc gctcctcaac ggcacgtcgc aactcgccga
 36151 cgaagagggc ctgttgttca ccggacggct gtcactgcag tcgcatccgt
 36201 ggctggccga ccacgccgtc atgggacagg tcctgctgcc cggcaccgca
 36251 ctgctggaac tcgcccgtcg cgcggcgac gaggtcggct gtgaccacgt
 30 36301 cgaggaaactg acgctcgccg caccgctcgt cctgcccag cgccgcccgc
 36351 tacagacca ggtccgggtc ggcgtcgccg acaccaccg gcgccgcacc
 36401 gtcacgatec actcgcgctc cgcacgcgc acgaccaccg acagtgcac
 36451 ccacaccggc accgacacc cgtggacca acacgccacc ggcgtcctcg
 36501 tcgcccgcct gccggcgacg gcaaccgtc cgttcgatgc caccgtgtgg
 35 36551 ccgccggcg acgccgaacc cgttgacctg gcggacttct acgcgtcccc
 36601 ggccggcgaa ggattcggct acgggcccgc tttccagggc ctgcgagcgg
 36651 cctggcgccg cgacggcgag gtgttcgcc atgtcgact gccggaggcc

36701 ggccgtaccg aagccgaggc gtacgggctg catccggcac tgctcgacgc
36751 cggactgcac gcagcctggc tcgtcgcccc ggacggggag cccacacgga
36801 cgggcagcgt gccgttctcg tggcgcggcg tcttcctggc cgcttccggt
36851 gcctcctcgg tccgcgtccg actcggccgc gactccgacg gaacgctgag
5 36901 cctggccatc gccgacacca ccggtgcacc ggtcgcgtcc gtacaggccc
36951 tctccatgcg caccgtctcg gtgacggccc tgagcgccac cgcgggcctc
37001 gcccgcgacg cgctgttccg cctggactgg gcctcggccc cggagccggc
37051 gtgccagccg gacgacacgg tgaccgtgat cccggcggtc gcggtcgtcg
37101 gtacggaaac ctccgaactc acctccgagc tcaccgcggc cctgcgtgcg
10 37151 gcgggcgccc acgtcgacgt ccgcacgacc ctgtcgaccg atgaaccgcg
37201 gcccgcgctg atcgcgctgc cgctcgtcgc ctccgaccag accggcaccg
37251 cagaagcggc acccgctccc gcggccgtgc acgacctcac ccgccgagcc
37301 ctggccctcg tacagacctg cctgcaagag cagcacttcg cggacacgaa
37351 gttcgtcttc gtgacctggt gtgcgacggc cgggcgtgat gtggctgctg
15 37401 ctgcggtgtg gggctctggt cgcttcggcg agtcggagaa tccgggttgt
37451 tttgctctgg tcgatctgga tccggatggt gcggtgggtg cggctgcgct
37501 cgctcgctcg ttggtcagtg gtgagccgca gcttgcggtg cgcggtgatg
37551 tgttgcggtg cgcgcgtctg gtgcggcggc cgctcaccga ggtcgggtcg
37601 ggtgctgatg gcaccgggga tggcgtcggg ggtggctctg gtgtgtcgtt
20 37651 ctcggtgtag ggtgcggtcc tggcactggt tggtaggggt ggtctgggtg
37701 cgggtgttggc gcgtcatctg gtggccgagt atgggggtcg ggatctgctg
37751 ttggtcagtc gcagtggta acgtgccgtg ggtgctgggg agttggtggc
37801 ggagcttgcg ggtgtgggtg cgcgggtcg ggtggttgcg tgtgatgtga
37851 ccgatcgtgc cgcggtggtg gagttggttg gcgggcatgc ggtgtccgcg
25 37901 gtggttcatg cggctggtgt gctggatgac ggcatggtgg gtgcgttgac
37951 cggggagcgg ttgtccgcg tgctgcggcc gaagggtgat gctgtctggc
38001 atctgcatga ggcgacctgc ggcctggatc tggacgcgtt cgtcgtcttc
38051 tcctctctcg ccggggtctt cggcagtcct gccaggcca actacgcggc
38101 cgcgaacgcc ttctggacg cgctgatgac gcggcgcccg gcggagggac
30 38151 tgcccgccct gtcactcgca tggggaccgt ggtcgtgac cgatggcaca
38201 tcgggcatgc tcgcgacgc cgaggccgat cgcctgacct gttcgggagt
38251 gccaccgctg accgcggagc aaggactggc actgttcgac gcggccctgg
38301 cgaccggtga cgccacctgc gtcccggtcc gcctggacct gtcggcactg
38351 cgtgcccagg gtgaggtgcc gcccttctgt cggctccctga tccgaggccg
35 38401 ctgcgcgcgc gccgcgcgc cggaatcggc aaccgccacc gggctgcggg
38451 aacggctcgt cggactgaac ccggtcgagc gacaggaagt cctcctggac
38501 ctcgtgcgcg gccaggtggc cctggtcctc ggccacgccc acgccgacga

38551 cgttcatccg gcacgtgcct tcagggagtt gggcttcgac tccctcacct
38601 cgggtggaact gcgcaaccgc ctcaacaccg tcaccggtct gcggctcccc
38651 gcgaccatgg tggttcgacta tccgaccgtc gaagtgtctg tctcctacgt
38701 cctggacgag ttgctgggca cggatgccga ggtggcgacc gtgcagccgg
5 38751 ccgcgggttg ggtggcggac gatccgatcg tcatcggtgg catggcctgc
38801 cgctaccccc gcgggggtcg ctctccggac gacctgtgga ggctcgtcac
38851 ggacggcgtc gacgcgggtg ccccttccc gaccaaccgt ggttgggacg
38901 tcgaatccct ctatcaccgc gaccccgacc atctcggtac ctctacacg
38951 cgctcgggtg ggttcctgca tgaggcgggg gagttcgatc cggggttctt
10 39001 cgggatgagt ccgcgggagg cgttggcgac cgattcccag cagcggttgt
39051 tgttggagtc gtcgtgggag gcgatcgagc gggccggtat tgatccggtg
39101 agtttgccgg gtagtcggac ggtgtgttc gcgggggtga tgtacagcga
39151 ttacagcgcg atgttggcga gtccggagtt caggggttcc cagggcagtg
39201 ggagttcgcc gagtttgccg tcgggtcggg tggcctacac gttggggttg
15 39251 gaaggcccg cggtgacggt ggatacggcg tggtcgtcgt cgttgggtggc
39301 gatgcactgg gcgatgcagg cgttgcgtag tggtagtgtt gggttggcgt
39351 tggccggtgg tgtgacgggt atgtcgacgc ctgcgggtgt tgtggacttt
39401 gctcggcagc ggggtttgtc gccggatggt cggtgcaagg cgtttgcgga
39451 tgcggccgat ggtgtgggct ggtccgaggc cgtcggcgtg ttggtcctgg
20 39501 agcggcagtc ggacgcgggt cgcaatggtc acgagatttt ggctgtgggtg
39551 cggggttcgg cggtaacca ggatggtgcg tccaatggtt tgacggcgcc
39601 caatggtccg tcgcagcagc ggtgatccg gcaggcgttg gccagtgggtg
39651 gtctgacggc cggtgacgtg gatgtggtgg aggcgcattg tacgggtacg
39701 acgctcgggtg atccgatcga ggcgcaggcg ttgttggcga cgtatgggcg
25 39751 ggatcgtgag cctgagcggc cgttgttgtt gggttcgggt aagtcgaatc
39801 tggggcatac gcaggctgct gcgggtgtgg cgggcgttat caagatgggtg
39851 ttggcgatgc ggcattggtg ggtgccgcgg acgttgcatg tggatgcgcc
39901 ttcttcgcat gtggactggt ccgaggggtg ggtggagctg ctcaagtgcg
39951 aggcggcttg gccggagacg ggtcgggtgc ggcgggcggg tgtctcctcc
30 40001 ttccggcatca gcggcaccaa cgcgcacgtg attctggagc gtccggaggc
40051 cgcgcggcgt ccggtgatgg agacgaacac cgtggagccg tccacggtgc
40101 cgtgggtgct ttccggcaag acaccggagg ccctgcgcgc ccaggcggcg
40151 aaactgttgt cgtcgatcga ggaacgccc gagcttcgcc tcgtcgacgt
40201 gggaatgtcc ctggtcaccg gccgctcgac ctttgaacac cgtgcgggtg
35 40251 tactcgccgc cgaccgtgcc gacgccgccc gcgcattgtc ggcgattgct
40301 gccgacgagg ccgatgctgc cgtgccacc ggccgcgtcg gcgcgggtcg
40351 tcacgcgggt ctgttctcgg gtcagggtgc tcaacgtctg ggcattggggc

40401 gtgagttgta cgagcggttc ccggtcttcg ccgaggtctc cgatgtcgtg
40451 gtcgaccacc tggacgccgc cttgcccgcc caggccgggt tgcgtgaggt
40501 gatgtggggc gacgatgcgg agttgctgaa cgagacgggt tggacgcagc
40551 cggctctgtt cgccatcgag gtggcgctgt tccggctggg ggagagttgg
5 40601 ggtgtccgtc cggacttcgt ggccgggtcat tccatcggtg agatcgcggc
40651 ggcgatgtc gccgggggtg tctcgctgga ggacgcctgc cgcctgggtg
40701 ccgccgtgc gacgtgatg caggcggttc cggccggcgg tgcgatgatc
40751 gcggtccagg cgaccgagga cgaagtcacc ccgcatctga ccgacgacgt
40801 agcgatcgcc gccatcaacg ggccgaacgc actggctgta tcgggtgtgg
10 40851 aggatgccgc cgtcgagatc ggggcgcggg tcgcggccga ggggcgtcgc
40901 acgacccgac tccatgtgtc gcatgcgttc cactcgccgt tgatggatcc
40951 gatgtggcg gagttccgtg tggtagcgga gggcctgtcc tatgtgtctc
41001 cgtccctccc cgtcgtctcg aatctgacgg gccagggtgg caccggccgac
41051 gaactgtgct cggccgagta ctgggtgcgc caggtccgcg aggcgggtccg
15 41101 cttegccgac ggggtgacgg ccctcgaagc cgagggcggt cggaccttc
41151 tggaaactcg cccggacggc gtcctcgccg ccatggcagg agcctcgctc
41201 accgaatcct ccctcgcggt accgtgtctc cgtaaggacc ggccggagga
41251 accggcgcca ctcccgccc tggcccagtt gcacatcgcc ggcgcgcgcg
41301 tcgattggcc cgtgtctctc gctggtgtgg gtgcggggcg ggtggagtgt
20 41351 ccgacgtatg cgttccagcg tgggtgggtc tggccgggtg gtcgggttgg
41401 tggttggtgt gatgtgggtg ctgtggggct tgggtctgcg gggcatccgt
41451 tggtgggtgc tgcggtggag ttggctgcgg gtgcgggggt ggtgttgacg
41501 ggtcgtctgt cgttgcgtc gcatgggttg ttggctgatc atgcgggtgat
41551 ggggcgggtg tttgttcctg gtacggcggt gctggagatg gtgatgcgtg
25 41601 ctggtgatga ggtggggtgt ggtcgtgttg aggagctgac gttggcggcg
41651 ccgttggtgt tgcctgagcg tggtaggggt cgggttcagg ttgctgtgga
41701 tgctcctgat gctgcgggtc gtcgtggtgt gggggtgtat tcgtgtcctg
41751 atggtgtggg tcaggcggtg tggtcgcagc atgctgtcgg tgtgttgcc
41801 tctggtgtgg ctgaccaggt cggtaggggt ggtgacgggt gtgtgtggcc
30 41851 gccgcagggt gcggtgtcgg tggatgctga gggctgctac gagctgtttg
41901 cggatgctgg gttcggttat ggcccgggtg tccaggggtt gcgtgcgggtg
41951 tggcgtcgcg gcgaggaact cttcgagag gtccgctgt cggacgaggt
42001 tgctgagagc gctgatacgg cgaccggttt cgggttcac ccggcgttgc
42051 tggatgcctc gctccatgcc tcgcttctct cctcccttga aggtcaatcc
35 42101 gccgatggtg ggcctgcgtt gccgttcgcg tgggaggggt tttccctctt
42151 cgctcgggt gcgacggctt tgcgcgtgcg gttggcgccg gcgggtgagc
42201 atgcggtgtc ggtgaccgcg gtggatccga ccggtgcgcc ggtgatttcc

42251 atcgacgcgc ttcgtacccg tcgcctcacc ctcgatgagg tcaacgcac
42301 tcacacccag ctgagcgatg cgtctctcgg cgtccaatgg accacgggcc
42351 cgagcacccc ggccgccgac caccgcgcgg tagccatcat cggaaccgac
42401 cacctggggc tcgccgaagc gctcagcagt tcctctgctg gcgcgaccac
5 42451 gaccacgacc gcggccgcgt acgagagcct tgacgcgctg atcgcggcgg
42501 ggcccgaagt gtccgtccct gacgtcacac tcatcggtct caccaccgag
42551 gacgccatcg ctgagtcagt gaacgaccac gacgccacgg tggccggcca
42601 aggcaccatc ggagccggcg cggcgccggt agacgcggct cgccgcctca
42651 ccgccgaggc cctgcgcacg atccaggcat ggttgggcca cgagcgtctg
10 42701 gcggcaaggc gcctggtctt cgtgaccgcg ggccggggcg acgggcagga
42751 tgtcgccgcg gcggccgctc agggcctggt gcgctccgcc cagaccgaga
42801 acccgggcac ctccggcctc ctgcacctcg acgggaccga ggcgtcgacc
42851 gcggtcctcg gcgaggtctt cactctcgac gaaccgcaac tgcttctcgg
42901 cgacggacac ctgcacgcgg ccaggctgac ccgctggcg tcgcccgcgg
15 42951 acacggccgt gccacggag tggaacgcgg acggcacggt gctgatcacc
43001 ggtggtaccg gcgggctcgg cgcgcagttc gcacggcacc tcgtcgacag
43051 gtacggcgtc cgcaatctcc tgcctgctcag ccggcgccgc cccgatgccc
43101 cgggaaccac ggagttggtc gccgagctga cggcgccacgg tgccgaggtg
43151 gccgtgcagg catgtgacgt ggccgatggc gatgcggtgg cggcggtggt
20 43201 cgccggcggtg ccggatgagc acccgctgag ggccggtcgt caccggcccg
43251 gtgtgttggc cgacggagtg atcggtctcg tcaccgagga gcggctcgcc
43301 accgtcctgc ggcccaaggc ggatgccgcc tggcatctgc acgaagcgac
43351 ccgcggcctg gacctggacg cgttcgtcgt cttctcctcc gtcgcagggg
43401 tcttcggcgg cgccggccag gccaaactac ccgcggccaa cgccttctg
25 43451 gatgcgctga tggcccagcg ccgggcagcg ggccctgccc gactctcggt
43501 ggccctggggg ccgtgggacc agaccggcgg aatgacgggc atgctgctcg
43551 acgcccaggc cgaccgcctc gcccgctcgg gcatcccgcc gctctccgcg
43601 gagcagggcc tcgcccctct cgacgcggca ctgctcttg ccggaaccag
43651 cacgccagac agggcagccg gcagcgccgc cgccagcacg tcgggaaccg
30 43701 gcgacacgat cgccatcccg gccgcggccc tcgtcgacac ggtccggctc
43751 gacctggcgg cgttgcccg gcagggtgag gtccccgcga ttctgcgcgg
43801 actggtgcgg acccgcaccc gccgtacggc ggccggcggt tccgtcaccc
43851 tggccggact cgtcaaccgc ctgtccgggc tcaccgccga cgagcggcgc
43901 caggaaactc tcgaactggt ccgcactcag gcagccctcg tcctcgggca
35 43951 cgccgatccg gcgtccgtgg actccaccgc acagttccgt gacctcggct
44001 tcgactcgct gaccgcccgc gagctgcgca accggctgag cacggccacc
44051 ggccctgcgc tgaccgcaac cctggtcttc gactaccgga acaccgatgc

44101 cctcgcggag cacctgcggg acgagttggt cggcgcggtc gagagcgagg
44151 tgcgggtgcc ggtccaggca ctgccgccga ccgccgacga tcccatcgtg
44201 gtggtgggca tggcctgccg tttccccggt ggtgtgacct cgccccgagga
44251 cctgtggcgc ctggtcgacg ccggcaccga cgccatcacc accttcccga
5 44301 ccaaccgcgg ctgggacctc gaatcgctct acgaccgcga ccccgcacac
44351 ctccggcacct cctacacccg ctccgggtggc ttcctgcacg aggcggggga
44401 gttcgacccg gcgttcttcg gaatgagccc gcgtgaggcg ctggcaaccg
44451 actcccaaca gcgtctcctg ctggaatcct cctgggaggc gatcgagcgg
44501 gccggcatcg acccgctgac cctgcgcggc agcgccaccg gcgtcttcgc
10 44551 cggcgtgatg tacagcgact acgggagcat cctcggcggc aaggagtctg
44601 agggcttcca aggccaggga agtgccggca gcgtggcctc gggccgcgtc
44651 tcctacaccc tcggcttcga aggtcccgcg gtcaccgtgg ataccgcctg
44701 ctccctctcc ctggtcgccc tgcacctggc agcccaggcc ctccgggcgg
44751 gtgagtgcac gctcgcgtc gccggtggtg tgacggtgat gtccacgccca
15 44801 ggcacgttcg tggagtctc gcggcagcgg ggtctggcgc ctgatggtcg
44851 ttccaaggcg ttcgccgagg ccgcggacgg cgtcggctgg tccgaggcg
44901 tcggcatcct cgtcctggag cgccagtcgg acgccgtgcg caacggccac
44951 gagatcctcg ccgtgatccg cggctcggcg gtcaaccagg acggtgcgtc
45001 caacggcctg accgcgccc acggcccgtc ccagcagcgc gtcacccgtc
20 45051 aggcgttggc cagtggcggc ctgtccacgg ccgacgtgga cgccgtcgag
45101 gcgcacggca cgggtacgac gctcggtgac ccgatcgagg cccaggcgct
45151 cctggccacc tacggccgcg accgcgaccc cgagaaccg ctgctgctcg
45201 gttcgatcaa gtccaacct gccacaccc aggcagcggc cgggtgctgc
45251 ggtgtcatca agatggatc ggcgatgcgg cacggcgtgc tgccgcagac
25 45301 cctgcatgtc gacgcgccgt cctcgcacgt cgattggagc gtcggcgccg
45351 tcgaactgct caccgagcag accgtctggc cggagaccgg ccgggtccgt
45401 cgcgcgggtg tctcctcctt cggcatcagc ggcaccaacg cccacgtcat
45451 cctggaacag ccggaggccg tgcagcgcct ggcaccggga gcagcagaga
45501 ccgtggagcc ggtcgccatc aagccgtcgg cggaaccgtc cctggtgccg
30 45551 tgggcgtgt cggcaagtc acccgaggcc ctgcgcgccc aggcgcacg
45601 cctccgtgac ttcctggcgg aacggcccga accgcgctcg atcgacatcg
45651 gccactcact ggccgtcaca cgctcgagc tcgaccaccg cgcgatcgtg
45701 ctggtcgacg atcgaaagg cccggcggac agcctggccg cctcgcggc
45751 cctggcctcc ggtgtggccg atcccgcgt cgtctccgac gcggtatcga
35 45801 ccggcggttc ggcagtgtg ttcacaggtc aggggtgtca acgtctgggc
45851 atggggcgtg agttgtacg ccgtttcccg gtcttcgccg aggtctctga
45901 tgtcgtggtc gaccacctgg acgccgcctt gcccgcaccg gccggtttgc

45951 gtgaggtgat gtggggcgac gatgtcgagt tgctgaacga gacgggttgg
46001 acccagcccc cgctcttcgc cgctcgaggtg gcgctgttcc ggctgggtgga
46051 gcgttggggt gtccgtccgg acttcgtggc cggtcattcc atcggtgaga
46101 tcgcggcggc gcatgtcgcc ggggtgttct cgctggagga cgcctgccgt
5 46151 ctggtggccg cccgtgcgac gctgatgcag gcgctgccga ccggcggcgc
46201 gatgatcgcg gtccaggcca ccgaggacga agtgaccccc cacctgaccg
46251 acgaggtggc gatcgcgcc gtcaacggcc cgacctccgt ggtgatctcg
46301 ggcgcagaag aggccacgca gaccgtggca caacacttcg ccgaccaggg
46351 gcgccggacg accgcgctgc ggggtctcgca tgcgttccac tcgccgtga
10 46401 tggatccgat gctggcggag ttccgtgcgg tggcggagg actgtcctac
46451 gccaccccgt ccctccccgt cgtctcgaat ctgacgggct ggctggccac
46501 ggccgacgaa ctgtgctcgg ccgagtactg ggtgcgccac gtccgcgagg
46551 cggtcgcgtt cgccgacggc atcaccaccc tcgaagccga gggcgtgcgg
46601 accttcctgg aactcggccc ggacggcatc ctgtccgcgc tggctcagca
15 46651 gtccctcgcc ggcgaagccg tcaccgtgcc cgctctgcgc aaggaccgcg
46701 gtgaggagtc cacggccctg acggcccag cgcatctcca caccgcgga
46751 ctgatcgaag actggcagga cttcttcgct ggtgtgggtg cggggcgggt
46801 ggagtgtccg acgtatgcgt tccagcgtgg gtggttctgg ccggttggtc
46851 gggttggtgt tgggtggtgat gtgggtgctg tggggcttgg gtctgcgggg
20 46901 catccgttgt tgggtgctgc ggtggagtgt gctgcgggtg cgggggtggt
46951 gttgacgggt cgtctgtcgt tgcgtcgca tgggtggtg gctgatcatg
47001 cggtgatggg gcgggtgttt gtcctggta cggcgttgct ggagatggtg
47051 atgcgtgctg gtgatgaggt ggggtgtggt cgtgttgagg aactgacgtt
47101 ggcggcgccc ttgggtgttc ctgagcgtgg tggggtgcgg gttcaggttg
25 47151 ctgtggatgc tcctgatgct gcgggtcgtc gtggtgtggg ggtgtattcg
47201 tgcctgatg gtgtgggtca ggcggtgtgg tcgcagcatg ctgtcgtgtg
47251 gttggcctct ggtgcggctg accaggtcgg tgggttcggt gacggtggtg
47301 tgtggccgcc gcagggtgcg gtgtcgggtg atgctgagg ctgctacgag
47351 ctgtttgcgg atgctgggtt cggttatggc ccggtgttcc aggggttgcg
30 47401 tgcggtgtgg cgctcgcgcg aggaactctt cgcagaggtc gccctgtcgg
47451 acgaggttgc tgagagcgt gatacggcga ccggtttcgg gttgaccccg
47501 gcgttgctgg atgcctcgt ccatgcctcg cttctctcct ccctgaagg
47551 tcaatccgcc gatggtgggc ctgcgttgcc gttcgcgtgg gaggtgttt
47601 ccctcttcgc ctcggtgcg acggctttgc gcgtgcggtt ggccggcg
35 47651 ggtgagcatg cgggtgcggt gaccgcggtg gatccgaccg gtgcgccggt
47701 gatttccatc gacgcgttc gtaccgctc cctcacctc gatgaggtca
47751 acgcatccca caccagctg agcgatgcgc tcttcggcgt ccaatggacc

47801 acggtcccgga gcaccccggc cgccgaccac ccgtcggtag ccatcatcgg
47851 aaccgacccc ttcggcctcg cagacggcct ttcggacgcc ttgcccctgg
47901 tcgaggagcg cggtagacctc gcggcgctcg cagcgtcgga gcacccggta
47951 ccggacctcg tcctcgctcc ggtagcgggc acccggcgca caggcgctacc
5 48001 tgcggacgcc gaaggacaca ccgacgccgg gacatccgac atgctccgat
48051 ccgtgcgtga ggccaccgca caggctactgg agcagatcca gcagtggttg
48101 gcggacgacc ggttcgaggc ggcgcggtcg gtgttcgtga cgcgcggggc
48151 ggtttccgtg ggtgaggggc gcatcgccga cctggcggcc tcggccgtct
48201 ggggtctggt gcggtcggcg cagtcggaga atccgggctg cttcgggtctt
10 48251 ctcgacctcg acctcgacct cgcccttgac tccgaccttg cccccgaggt
48301 cgacatcgag cgcgaccgtg acccgcatcc ggtcgggtgg accgtgcagc
48351 ccgcgctcgc cgcggccctg cagcgacccg ccgacgagcc gcagttggca
48401 ctgcgcggcg ggaccgtgca ggccgcccga ctgacccgaa tccccgcgcc
48451 gcagaccgac cgtgccgaga ccgacctgc cgagaccgac cgtccggaga
15 48501 tcgacacccg gcggcccggc acggtgctca tcaccggtgg taccggtggc
48551 ctcggtggtg tgctcgcccc gcacctcgtc gccgagcggg ggggtacggag
48601 cctggtgctc gccagccgga gcggtctcgc ggccgagggg gcggagaagc
48651 tggtcgccga cctcgaagcg ctcggtgccg tggtgccgt gcagacgtgt
48701 gatgtggccg atggcgatgc ggtggcgggc ttggtcgccg gcgtgtcgga
20 48751 cgagtacccg ctgacggcgg tcgtccacac ggccgggtgtg ttggacgacg
48801 gagtgatcgg ctcgctcacc gaggagcggc tcgccaccgt cctgcggccc
48851 aaggcggtg ccgcctggca tctgcacgag gcgaccccg atctggacct
48901 ggacgcgttc gtcgtcttct cctccctcgc cggcgctctc ggtggcgccc
48951 gtcaggccaa ctacgcggcg gcgaacacgt tcctggacgc cttgatggcg
25 49001 cagcgtcgcg ccgccgggtt gccgggtgtg tcgctggcgt ggggtccgtg
49051 ggaccggggc ggcggcattga cggggaccct gtcggacgcc gaggccgacc
49101 gcctcgcccg ctccgggtgt ccgccgatct cggcgagca gggccttgcg
49151 ctgtacgacg cggcgaccgc cggtagcgg ccgctggtgg tgccggtgcg
49201 gctggacctc gccgcgtcc gcgggctcgg tgatgtcccg gcgctgctgc
30 49251 gcggactggt ccggacgccc gcgcggcgga ccgcggcggc cggtagcgcg
49301 ccgtcggccg atgtgctcac ccggcagttg gccgggctcg gcggggcgga
49351 gcaggaggag gtcctgctga ggctggtgcg cggtcaggcc gcggtggtgc
49401 tcgggcacgc cgacggctcg gcgatcgggt cggggcgaca gttccaggag
49451 ttgggcttcg actcgctgac cgcggtggag ttccgcaacc gactcaacgc
35 49501 ggccaccgga ctgcggctgc cggccaccct gctgttcgac taccgacgc
49551 cggccgacgt cgtcgggcac ctgcgcggcc ggctcggcac cggggagggtg
49601 tcgggtgcgg gctcgggtgt ggcgcgctg gacaaccttg aggcggtgat

49651 cgccgggctg tccctcgacg acgcggggga gcaccagttg gtggccggcc
49701 ggctggaggt cctcagggcg aagtgggchg acatgcgaag cgcgagggga
49751 gctgtggacg gcggtgcgga cgtcgacatc gaggaggcgt cggacgacga
49801 catgttcgch ctgctggacg acgagctggg gctgaactga gccgctccgc
5 49851 atgagcagtt ccgcaccaga agttccacag tggcctggtc cacagtgacc
49901 tgggtccacag tgagctgttc cacagcgacc tgttccgchg cggttccgta
49951 gtgagccggt ccgcttaggc gggaaaccgg gggcccggtg gtcggacgct
50001 tgattccggc caccgggagc gtcaaaccgg cctcttctaa agaaagggaa
50051 agaactaggg aaacacccac agcccgatca tgtgcaatga attccgtggc
10 50101 tcggaaaacc attcgcgggc catggagtta cgggtgtatc acgtgtccgt
50151 acggcagtga cggcgtagc gggttgccac ggcagggtccc cggccgaggc
50201 gtctccccgc tgcgtccchg cggtagctgc cgcttgggtc gtccctagga
50251 ggtcccccca tgaccacgtc caccgaggag agcctgtggg cccggtgctt
50301 ccattccagca ccggccgccc ccgtccggct cttctgtttc ccacatgcgg
15 50351 gcggctcggc ctccctctac ttcccggtgt cggcccaact gtctcgggt
50401 gccgaggtgt tcgcatcca gtaccgggg cgccaggacc ggcgaagga
50451 agccggtgtc agtgacctc cgaccttggc cgaccaggtc tacgacgcgc
50501 tgcgccccct gctgaaggag cggccgagca cgttcttcgg gcacagcatg
50551 ggcgcgacgc tggccttcga ggtggcccgg cgcttcgagg ccgacgacgg
20 50601 tgacctggtc cggctgttcg cctccgggch ccgggcccc tcccgctgc
50651 gtgaagaggc cgtgcaccgg cggtagcagc acggcatcgt cgaggagctg
50701 aagctgctc cggcaccaa caccgcgctg ctcggcgacg aggagatcct
50751 gcggatgac ctgcccgcga tccgcagcga ctaccaggcc atcgagacct
50801 accgctgccc gcccgacgtc accgtccggg cgccgctgac cgtcctcacc
25 50851 ggcgaccgcg acccgaagac ctccctggac gaggccgagg cgtggcgchg
50901 ccacaccacc ggggacttcg acctcaagg gtctcccggt gggcacttct
50951 tcgtcagctc cgaggccccg gcgatcatc atctgtctcg ggcgcacctc
51001 gccggcaacg gctagcgggc gactgcggc aggcggcgcg tgccggtctg
51051 ccgcacggcc cccgcgcgc ctgagacggc accatgccac ccggcgacgc
30 51101 gcgttgchg cgctgtgga gcaccgtccc gcgtgcgtac gcgggcgcgc
51151 ggaggacgtg ctcgcccggc cggatgtcct ggaggagacc attcgaaaa
51201 cgtcaaatgt ctgacagctg ctgctcatct actccagccc gcactatatg
51251 atcttgtggt ggggcatgta cggagcgtgc cactcggttc ctctggtgc
51301 atccatccc accgtccgga tttgtgtgca ccggacggcc gtggtcctgg
35 51351 tacgcgchg cgcgaggtag gctcctgttc tccccacg tggaggaacg
51401 cgtgatcgcg aagcagtcgg gttcttcchg cttactgacc acgctggtgg
51451 gacgcgacga cgaactgcgc accctcgccc ggcacgccc ggcgccccg

51501 gacgggaggg cgggcctggt cctgctccac gggcccgccg gcatgggcaa
51551 gaccagtctg ctgcggtcct tcacggcgag cgatgtctgc cgcggcatga
51601 cgggtgctgta cggcacctgc ggcgagaccg tcgccggcgc cgggtacggc
51651 ggtgtgcgcg aactcctcgg cgggctcggc ctgagcggcg gcgacgcccg
5 51701 gcgctcgccc ctcttgagg gcctggcggc ccgcgcgctg cccgcactca
51751 ccgcagaccc cgcgggtccc gacgcgcga cgggtgccta cccggtgctg
51801 cacggcctgt actggctcgc cgcgcgcctc atggcccaac ggccgctggt
51851 cctcgtcctc gacgacgtcc actggtgcga cgaacgctcc ctggcctgga
51901 tcgacttcct gctgcgacgc gccgaggacc tgccgctgct ggtcgtcctg
10 51951 gcctggcgca gcgagggcga accggtcgcg ccgcgggtgc tcgccgacat
52001 cgcgcgccag cgcgcgcga ccgtgctcgg cctgcacccg ctgcgccccg
52051 acgacatagg cgaaatggtg cgtcgcgtct tccggaccac ggccgcaccg
52101 tcgttcgtca gccgggtcgc cgcgtgtcc gccggcaatc cgtggccct
52151 cgcgcgcctc ctgcagaaac tccgcgcga gggcgtccgg ccggacgccc
15 52201 cgggggagcg cggggccgccc gaggtcgga gtcacgtcct cgcgcgctcg
52251 gtgcgctgcc tgttgagcg cgggcgcgc tggtgctgcg gcgtggcccg
52301 tgccatcgcc gtactcgccc cggagtgcac cgagttgctg gcggcgctcg
52351 cggcgctccc ggccgcgacc gtcgacgagg ccctgttggg gctgcgcagg
52401 gccggcatcc tggccgcga ccgcgtggac ttcgtccatg acgtcgtccg
20 52451 ctccgcggtg ctgcagacg tcgccccgc caccctggcc gaactgcgca
52501 ccaacgccgc gctgttctg agcgacgccc gccgcgcctc cgaggagctc
52551 gccggccagc tcatgctgct gccggtgctc gaccagccgt ggatggccgc
52601 ggtgctgcgc gacgcgcgcg ccagggcga gagccgcggc gccccggagg
52651 cgggtgtgcg ctgcctctac cgggtgttg aggtggagcc ggacaacgtt
25 52701 gccgtccgca tccagatggc ccgcgcgctc gccgagatca acccgccga
52751 ggcgatgcgc ctctcaagg aagcgtctc cctcgccggg gacgtccgca
52801 cccgcgcga ggtcgcgctc cagtacggt tcacctgct cgcgtgcag
52851 gaatccccgt cgggggtgcg gatgctggag gacgcgctc ccgagctgac
52901 cgcgaactg ggcgcgaac caggggccgt ggaccgggag ttgcggaccc
30 52951 tcgtggagtc cgtgctgctc atcgtcgggg ccgacgagaa ggtgacgatc
53001 ggtgcggtcc gcgaccgggc ggcgcgctc accatgccgc ccggcgacac
53051 accggcccag cggcagatgc tggccatgac caccgtgctg accgcgatgg
53101 acggccggga cgcgcgctc gccgtcgacc aggcgcgcg cgcctcgcg
53151 gcaccggcg tcgagctgga accctggtc ctgctgtcc cctccttcgc
35 53201 cctctccctg gccgacgagg tcgccgacgc gcagtacgca ctggacctca
53251 tgctccagta cggccaggac aacgcggcgg tgtggacgta cgtcctggcg
53301 ctctccacgc gcgcctgct ccaccacggg gtgggcgcct tccccgaggc

53351 cctggccgac gcccagaccg ccgtcgagat cctcggcgag gagcgctggg
53401 cggacggcgc cgtgctgccc cgtgtcgcg cgtgtcgcg tggccaccgc cctcgtcgac
53451 cgcggcgagc ccgagcgcg cgaacacgtc ctcgacggca tcacccgccc
53501 ccgcctcgaa cgcttcgtca tcgaatacca ctggtacctc caggcccgcg
5 53551 cctacgcccc ctgggtccgc ggggatttcc aaggagccct ggacctcctt
53601 ctcgcctgcg gtcggtccct ggaggagtgc cgcttcagca acccgcgctt
53651 cgtgccgtgg tgggccgatg gcgcggtgct cctggcgacc ctggaccgcc
53701 acgaccaggc gcgcgaactc gccgcatacg gaagcgagtt ggccgagcgt
53751 tgggggacgg cgcgcggcct cggactggcc ttcattggccc agggcgctgc
10 53801 cgcacccggc cgcgcggcga tcgatcacct caccgaggcg gtctcgctgc
53851 tcgcggaact cccggccccg gccatggagg cccgggcccga acttcttctc
53901 ggacacgccc acctgaagcg cgacgacctg cgggcccgcc gggaacacct
53951 gcgcgccgcc gccgacctgc cccagcgctg cggcgccgtg aagctcggcg
54001 tcgacgccag aaaactgctg gtcaccgcgg gtggtcgggt acgcaggatg
15 54051 accgcctccc cactcgacat gctgaccggg atggaacgca cgggtggcga
54101 cctcgcggtg accggcgcg gcaaccgggc catcgcgga gccctcttcg
54151 tgactgtaag gaccatcgaa acccatctca cgagcgtcta ccggaagctc
54201 ggggtcggcg ggcgtgcgga gctgtccgcc gtcttgaga ccaggaccgc
54251 cacctccggt cggcagccgc cggcctgggt ctcccaggca cgcggacgcg
20 54301 cttgagacga acaggacgag aggtagcggg gccccggagc aaggcgcgga
54351 accaaccgac cacctgcacg ccgcagtgcg caccgacgc ccacggcgac
54401 cccaccatgc tcctcgaatg cggccgggaa cagcggctca tcggcgacct
54451 cctgcaccgc ctcggccagg ggcggccatc ggtgctcagc ctgaccggcc
54501 gggccgggca cggccagaac gccctggtcc gctggggcgc gtgcccggcc
25 54551 aggcacgacg ggctgcgct cctgcgagcc caggcgacgc ccgcggaacg
54601 ggaactccgc tacggcgccg ttctccaact gctggccgtc ctcgacggcc
54651 cgcacggcag caccctggac gccgcgatcc gccacgacgg tcccccgcca
54701 ctgcccgtgc ccggcatcga ggaggtgctg cggcgacacc gcacggcacc
54751 caccctggtc gtggtcgaag acgtccagt gttggacccg gcctcgctga
30 54801 cgtggttgca gatcctgctg cgccacctgc ggccggacac cccgctcgcc
54851 gtcctggcca gcagctgcgg tgacaccacg gccttcgaca ccgacccgaa
54901 gggccccgcc gtcccggggc cgcgggacac cgtgcccgtc gcgcgcttcg
54951 tgggtcccg cgtcaccgac cgcggggtcg ccgccaccgt ccgcgccgtc
55001 tgcggcacc cggcgacga ggagttcatc gccgcgtca cctccgccac
35 55051 cgcggcaac cccgcatcc tgcgggacgc cctgcgcgcc ttcgtcgacc
55101 acggcctccc cgcgcacgc gaccacctcc cggagctgca cgccctcacc
55151 gctggcgctc tcggcgacca caccgtgcgc gccctggacg gcctgcccgc

53351 cctggccgac gccagaccg ccgtcgagat cctcggcgag gagcgctggg
53401 cggacggcgc cgtgctgcc cgtgtcgcgc tggccaccgc cctcgtcgac
53451 cgcggcgagc ccgagcgcgc cgaacacgtc ctcgacggca tcacccgccc
53501 ccgcctcgaa cgcttcgtca tcgaatacca ctggtacctc caggcccgcg
5 53551 cctacgcccg ctgggtccgc ggggatttcc aaggagccct ggacctcctt
53601 ctcgcctgcg gtcggtccct ggaggagtgc cgcttcagca acccggcgtt
53651 cgtgccgtgg tgggccgatg gcgcggtgct cctggcgacc ctggaccgcc
53701 acgaccaggc gcgcgaactc gccgcatacg gaagcgagtt ggccgagcgt
53751 tgggggacgg cgcgcggcct cggactggcc ttcattggccc agggcgctgc
10 53801 cgcacccggc cgcgcgggca tcgatcacct caccgaggcg gtctcgctgc
53851 tcgcggactc cccggcccgg gccatggagg cccggggcca atttcttctc
53901 ggacacgccc acctgaagcg cgacgacctg cgggcccgcc gggaacacct
53951 gcgcgccgcc gccgacctcg cccagcgctg cggcgccgtg aagctcggcg
54001 tcgacgccag aaaactgctg gtcaccgcgg gtggtcgggt acgcaggatg
15 54051 accgcctccc cactcgacat gctgaccggg atggaacgca cgggtggcgga
54101 cctcgcggtg accggcgcca gcaaccgggc catcgcggaa gccctcttcg
54151 tgactgtaag gaccatcgaa acccatctca cgagcgtcta ccggaagctc
54201 ggggtcggcg ggcgtgcgga gctgtccgcc gtcctggaga ccaggaccgc
54251 cacctccggt cggcagccgc cggcctgggt ctcccaggca cgcggacgcg
20 54301 cttgagacga acaggacgag aggtagcggg gccccggagc aaggcgcgga
54351 accaaccgac cacctgcacg ccgcagtgcg caccgcacgc ccacggcgac
54401 cccaccatgc tcctcgaatg cggccgggaa cagcggctca tcggcgacct
54451 cctgcaccgc ctcggccagg ggcggccatc ggtgctcagc ctgaccggcc
54501 ggccccggga cgcgcagaac gccctggtcc gctggggcgc gtgccccggc
25 54551 aggcacgacg ggctgcgct cctgcgagcc caggcgacgc ccgcggaacg
54601 ggaactccgc tacggcgccg ttctccaact gctggccgct ctcgacggcc
54651 cgcacggcag caccctggac gccgcgatcc gccacgacgg tcccccgcca
54701 ctgcccgtgc ccggcatcga ggaggtgctg cggcgacccg gcacggcacc
54751 caccctggtc gtggtcgaag acgtccagtg gttggaccgg gcctcgctga
30 54801 cgtggttgca gatcctgctg cgccacctcg ggccggacac cccgctcgcc
54851 gtcttgcca gcagctgcgg tgacaccacg gccttcgaca ccgacccgaa
54901 ggcccccgcc gtccccgggc cgcgggacac cgtgcccgtc gcgcgcttcg
54951 tgggtgcccgc gctcaccgac cgcggggctc ccgccaccgt ccgcgcgctc
55001 tgccgcaccc ccggcgacga ggagttcatc gccgcgctca cctccgccac
35 55051 cgcgggcaac cccgccatcc tgccggacgc cctgcgcgcc ttcgtcgacc
55101 acggcctccc cgcgcacgcc gaccacctcc cggagctgca cgcctcacc
55151 gctggcgctc tcggcgacca caccgtgcgc gccctggacg gcctgcccgc

55201 cgaagtcaac gccgtcctgc gggccctggc cgtctgcggc gacctgctcg
55251 acttccaccg agtccggggc ctcgccggcg cgcactcgct gtccgaggac
55301 cggatccgca ccctgctggc gagcgtcggc ctgaccgtgt ccgtcggcga
55351 caaggtgcac atccgcttcc ccgcctccaa ggcacgcgtc atcgaggaca
5 55401 tgcccgcgcg ggagcgcgcg gatctgtacg tccgcgcggc cgaactcacc
55451 cacagttgcg gcgtcaacga cgaggacgtc gcccatctgc tgctgcgctc
55501 gtcgccgctc ggcgcaccct gggtcgtgcc cctgctccgc cgcggattcg
55551 ccgccgcgct gcgccgggag gaccaccacc gggcctgtgc ctgcctctcc
55601 cgcgcctgc aggaacccct cgaccccccg gaacgcagcc tgctgacgtt
10 55651 ggaactggcc gcggccgaag ccgtcgcccc gccggaggcg ggggatcgac
55701 gcctggggga actcgtccgc agcaccgtcg cggacaccga cccacgtcg
55751 tccgtgaggg ggggtgggggt ccgcgccatc gacctgggggt tcgcccgggg
55801 caacagcgaa tgggtccgcc gcaccgcggg cgaggccctg ccgtacgccg
55851 ggccggccga ccgggaggaa ctggtcgcgc tgttctgggt ggccgccgtg
15 55901 cgggacgacg acgcgccgat gatccccgtg gtgccccggg tgcccgaccg
55951 gccggtgccg ccggcccagg ccggcgcccc tgccctggcag ctggccacgg
56001 cgggggagga cgcggacaag gccaggaagc tcgccgggat cgccctcacc
56051 ggcggggtga acgagagcct gatgatgccg aaactggcgg cctgcgcgcg
56101 gctgttcgcc accgacgaca acgacgaggc ggtgcacggc ctggacacca
20 56151 tgctcaccgc cgcgccaggt gccacactgc gcagcatggc cgcccgcat
56201 ttcaacctac gggcccggat acacctgtgc gcggcccggc tggaggccgc
56251 cgaacgcgat ctggacagcg ccgagcgcgc cctgccgcgg acgagttggc
56301 acccccgtgc gctgcccac ctgatcgcca cccgcatcct cgtcagcatg
56351 gagacgggcc gcccggaacc gggccgccga ctgcgcgagg ccccggtccc
25 56401 cgccggcggc gaggaggggt tgtggtggcc cgccctgctg ctgcgccgcg
56451 cccgtgtggc cgcgcacgac ggtgactggg aggaggccct gcggtgtcg
56501 cgggagtgcg ggcgctggct ttttcgccgg cactgggcca acccgccat
56551 gctcagttgg cgcgcgtgg ccgccgagg gtgtctgaag ctcggcgacg
56601 tgacggaggc gcgccggctg cgggacgagg agctgttctt cgccgaccgc
30 56651 tggggcaccg cgagcgcgcg cgggatcgcc cgctgacga cgcggcgact
56701 cttcgacgac gacggcgacc gggccgtccg gcggatccgc gaggccgccg
56751 ccctgctccg cgactcgccc gcccgctgg cctacctgtg gagccggctg
56801 agccaggccg gtgccgagac gggccacggc gacaccgccg cggccgcacg
56851 ctctggcag gcggtcgccc ggatgaccgc cgccacccc gccagccgcc
35 56901 tcgccaccgc cgcgcgacc ctgaccgtcc cgtccgttcc ggtcgccacc
56951 gcgccgcca ccgccgtcg cccaccgga tggcgcgacc tgtccgaggc
57001 ggagaaggac accgtgctgc tcgccgcccg cggccacggc aaccgccaga

57051 tcgccgaaca actcgccgctc agcaggcgca ccgtggagct ccggctgagc
57101 aacgcctacc gcaagctgag gatcggcgga cgcaaggagc tgtacctgct
57151 cctggaggcg ctggaaggac cggtcgcgga tgcttcttga gcgggagaac
57201 gaactggccc ggatccgggc cgccctggac gccgcggaag cgggcgactc
5 57251 ctcgctcctc ctgatcaacg gtccctcctcg cagcgggctg tcggcgctgc
57301 tgcgccgat accggagctg gccggcgacg gcacccgctg cctgccccgc
57351 agcgccgctt ggccgggaac cgacttcccc ttcgggatcg cccgccaaact
57401 cttcgaccac ctgctgtccg gggcgggcg cgcaggggcg gccgaacgca
57451 ccgccggggc agagcacttc agccgactga tggacaccgg cgaccgccct
10 57501 accgggaccg gccccgcctt ggaggtctcc caggcagtgc tccaggcgcg
57551 ccaggcgctg ctcgccgacg cgtccgcgga gcggcgctg ctgatcctgg
57601 tcgacgacct ccagtgggac gacggcccg ccctgcgctg gctggccac
57651 ctcacccggc ggctgcacgg cctgccccgc ctgctggtgt gcacgctggc
57701 cgacggcgac caccggggca ggtacccctt ggtccgggag gtcgccggcg
15 57751 ccgcgcacac cgtcctgcgc ctggcgccgc tgtcccgga cgccaccgc
57801 gtcctgctcg ccgggccccca gggccggcg ccgcaggacg cactggtgcg
57851 cgccgtgtac gagcggtcca ggggcaacc gctgttctct accgccttcc
57901 ggagcgctct cgcgccacc ggaaggccgc ccggcgcgga ccacttcggc
57951 gccgtccggg agctgagccc gacggtgctg cgcgatcggc tcgccccca
20 58001 tctgcggatc cagccgcagc cgggtgcgca ggtcgcggtg gcgggtggcg
58051 cgctgggga ccacagcgat ccggtgctgc tcgccagct cgccggggtc
58101 gatgaaatcg gtttcgccg tgcccgcgc gcgctggtgg acgccggcct
58151 gttggcccg ggacgggacg tccgcttctt ccacggcgct gtcgcgatg
58201 cgggtggact cctgctcacc ctcgacgagc gggaacgctc gcacgacgac
25 58251 gccgccgatc tgctgtaccg ctgccccg cgggccgagc aggtcgccgg
58301 ccactgctg gccgtggtcc acccgggcg gccctggtcg gagcggttcc
58351 tgcgctccgc ggcccaaac gcgctgcgc ccggccggcc cgccgacgcg
58401 gcccggtacc tgcgcgcgc cctgctgcac caccgcacc aggacggctg
58451 ccgcgccgc atcctggtcg atctggccac cgccgagcg gccctcgacc
30 58501 ccgatgcctg tgtacgccac gtcagccagg cggtcgcgct gctggacacc
58551 tcgccccgat gggccgcgc cgtgttgcgc atcccgcgt ccctgctcgc
58601 cgccccagc ccgtccgcg tcgagttggt gcggcaggcc gccgcgggc
58651 tcgacgaacc ggggcagcg gacgaggagg gagccgacga actcgccctg
58701 cgcttgagg cgtggctgcg gcactccggc cagagaacc ccgtcgagct
35 58751 ggcgtcctcg gtggcgcggc tgcggcgcat gggggcacgg ccgccggtgg
58801 acagcgtcgc cgaacgcgag ttggtcgcc tgctgttgag tgcggcgcg
58851 ctcagcgcc ggctcagcg cgccgagatc gccgacacc gcaaccgcat

58901 cctcgaacgt gagccggcca ccgccgcca tgcccacacc ccgctgccgc
58951 tggatgatgct ctcgctgttc gtcgccgagt ccgtgcaggg cgtggcctcc
59001 tggctggcca gcgaacagca caccggcgcc cggtagcga ccggcgcgga
59051 cgatgtgctg ctgaccgccg agcgggcctt cgtcctgggt acgcagggcc
5 59101 gcccggccgc cgcgcgggag cacgtcgaac gcgccctggt catggacgcc
59151 ggcgactggt cggaaaccgc cgtcatgatg ttcgccgcgg tcgccttcga
59201 actgcgcgac ccggccttga gcgaacgcat cctggaacgg atccgcgacc
59251 gccggccggc cggactggcg ctcaccgcca ccggtcagat gctccaggcc
59301 gccgtcgacg tgcacttcgg ccggggcgcg gacgccctgg acacgctgct
10 59351 ggcctgtggc cgacgcctgg agaccgtggg atggcgcaat tccgcactgc
59401 tgccctggcg tccgtatgca atcgggctgc accagcggct cggcgagacc
59451 gatgccgcgc tgcaactcgc cgaggacgag ctgaggtggg cccgggagtg
59501 ggggtgcgacg acgaacctgg gccgggcctt gcgcctgaag ggctggctgt
59551 tgcaggacga gggactggat ctgctgcgtg agagcgtcga gatcctccgc
15 59601 gcttcgtcct acgcgacgga actggcgcg accctcgtcg tcctcggggc
59651 tcggctgccg ggtgggcccg aggccgaggg ggtgttgccg gaggccgcgg
59701 ggatcgccgc cgcctgtggt gttccctggc tcgccgaacg ggccgaactc
59751 ggtctgggca gtgccatcgt gccgccggtc gccaccctga cccccagcga
59801 gcggcgagtg gcgtcgctgg tgagccgggg cctgaccaac caggccatcg
20 59851 cgaccgaact cgggtgtgag tcccgggcgg tggagaagca cctcaccagc
59901 gcctatcgca agctcggcgt ctcgggcgc cgcgagttgg tgaatgcgtc
59951 cccgggtcgt tgacgcccgc ggccgtccc ttcgggacga tcacaatcct
60001 ctgtgcctgt cagccctgtg ggtgcggcgt cgttgctgta cgcgcgtccg
60051 gtcattgacg caatcctcaa caaaagctca aacggaagcg ggggcagggtg
25 60101 catccacccc atcaccaatg tctcgtatgt tacttttggc ccaccgtagt
60151 tgggactctg cggaccgtac ctcgtcaagt acggaccacg gatttagggg
60201 tgcgtagatg tgtcgcgttg acgtgccgag ttgtggctaa ctacgctccg
60251 tttcacggaa atcgaaactc ttcggcgtga atcgcgcgga acttcgatcg
60301 aattctggat tccgtttctc agaaggctaa aagacgacgg gggttatccg
30 60351 aaaggatggc ggtctcgtga ctatcactca cctcactgac tcgaatcacc
60401 gtacgaccgg cggggtgata tccgctcaga ccgcaccggc cggggagtc
60451 gtcggccccg ggctgatggc gtccctcgac cgtgacctca ccatcaagca
60501 cgccaaccag gagttccgcc gccgcttcga cgattccgcc ggtgacgtct
60551 gcggtcgcag cttccgggac ctgatgcacc cgagcgtgca gcagccgctg
35 60601 atgcgccagt tctcccggct gatcgagggc aagcggcacc gcttcgcctc
60651 gcacgtggtc gccgtgggcg cccaggacgc ggccttcgcg ggcaccctca
60701 ccgcctccgc ggtcaccggc aagaccccg acatcgccgg gatcctgggtg

60751 ctcattggact cgtccggcgc cgccgacgcc gccgatgccg gcgtcgtgac
60801 cagccagaag aagttcctga ccgagatcga tgcgcgcata ctcgaaggca
60851 tcgcggcccg gctgtcgacc attccgctcg cctcgcggct ctacctcagc
60901 cgccagggcg tggagtacca cgtcaccggg ctgctgcgga agttgcgggt
5 60951 gccaaccgc gccgcgctcg tctcgcgcgc ctactccatg ggcatacctga
61001 acgtcggcac ctggcccccg aaggtcgtgg acgacttcat caagtgcgc
61051 ccggccggcc tgtcccgcac cgcccgccca ccgtcggacc gctgtcccg
61101 acgcccgtgt cgccgacgcc cgctctcac ggggcgatgc ccggcggcac
61151 gggcgtttcg tgccctggtc gatgtttatc cattctcacg ttacgaaccc
10 61201 gcaatgctca tgcgcgcgat tccggccact tttcgttccg tcgtaaccaa
61251 actgcccgtc atatcgtttc cactgaccgt ttctgtggtc cgcggtgtgc
61301 cacctgggccc ccactccgac gcgcccgcga cgggtcgcga gcggcctccg
61351 aatactcacg gtaacccccg ttctcctcgtg tggagtcccc cgcgcgtcac
61401 attcctccgt gcctcgtgcg gcgtggctcg cccgctttcc ccgcccgcgc
15 61451 gacagccatt gttacaacgg cagcagagag cgggggtggt cgaaccgtcc
61501 gcggtgcggc gcagtcaagc ggaacctctg ctcgatgttc gtgccacccc
61551 gaactctgtg ttcttggtga ccgttttcac caatgccccg cccgttcgggt
61601 ggcgaatccg ccgcccgttc ctggccacaa cggatcgctt cggccgcccc
61651 ccacggccgg cgagccgag gcttggtgga atcaccccca ccagtcggg
20 61701 gcgcccgtgt ggcgggaca cgcaggagga acaccgtgga tgctgagggc
61751 cgccgcaggg acatgctgga gctgatccgg cgcagcggat ccgcccgtgt
61801 cgtgcgtctc gccgaggagt tcgcccgcag caaggaaacg gtccgcggg
61851 acctgaacgt cctggagggc catgggctga tacggcggcg gcacgggtggc
61901 gcctatccca tgggtgcggc gggctccgag gccgtcttcg tctcccggac
25 61951 cgcacagccg atcccggagg agtcccggat cgccaccgcg gccgccgaac
62001 tcctcagcga ggccgagacg gtcttcatcg acgagggtt caccgccaa
62051 ctcacgcgcg acgcccgtgc gcgcgaccgg ccgctgacca tcgtcaccgc
62101 ctcgctcccg gtggtcagcg ccttcgcgac gagccacag gccaacgtgc
62151 tgctcctggg cgcccgggtc cgccggggca cgacggccac cgtcgaccac
30 62201 tgggcccgtc acatgctgtc cggttcgtc atcgacctgg ccttctcgtg
62251 cgcggagggg atctcgcgca ggtacggcct gaccaccccc gaccggcgg
62301 tcgcccaggt caaggcccg gccatccgcg tcgcgcgccg cccggtcctc
62351 gccggggtgc acaccaagtt cggcacggcg agcttctgcc ggttcggaga
62401 ggtgggagac ctggagacga tcgtcaccgg cgccggcctg cccgtcgcg
35 62451 aggccaccg ctaccacctc atgggcccc aagttttacg ggtgtgacgc
62501 cgccggggcg tcccgcgcgc cgccacggc ctggcgcgcg ccggtgcggg
62551 tcagccgcgt cgcgcgtccg tcggttcccg ccagagtgcg agcgggtgtca

	62601	ccccacccg	gtagccctcc	tcgtcgagga	gatggccgcc	gtgcttcacg
	62651	tgccagaggg	cgaccgcgcg	ttcgggtgacg	gtcaggtccg	cggcccgttc
	62701	ggcgagtcgg	acctcgaagc	gctgggctgc	gtcgagggag	cgcacccagg
	62751	ccaccaggtg	caggttgtgc	cggtgatga	cgctcgcgca	caggcggacc
5	62801	tcgcgcatcc	cggtgacccg	gcgggtcacc	tcgcgcagcc	tcgcggccgg
	62851	gacctgcccc	cagaagctca	ccgtgaccgg	ccactcggac	agcgggaggg
	62901	cgacctcgca	ccgggctgct	aacatgtccg	cggcgaagag	ccgttggacg
	62951	cgacggcgca	cggtgtcggg	gccggcgccg	cactgctcgg	ccagcgcgcg
	63001	gtaggtggcc	cggccgtcca	cggacagcgc	ggtgaccagc	tgttggctga
10	63051	gctcgtcgag	gacgaaggcg	ggggtgtccg	tgcggtgtct	ggacgcgtcg
	63101	gcggcgagcc	ggcgacctg	gtgccggccg	agggcccgca	ggcgccaccg
	63151	gctgccctcg	gtgtgcaccg	ggccggccag	gtgcgtccgg	gcggcccggg
	63201	cgccgtccaa	cgcggcgagg	tcgtgggtca	cccaacggga	gagcatggcc
	63251	ggatcgcggg	ccatgacgtt	gagttggagg	tcacgggtccc	cggtgacgtg
15	63301	ggagagcgcc	accacgtgcg	gcgcggcggc	caacgcccgc	gcgacgtcga
	63351	gcagtcggcc	gggagcgag	tcgatctcga	cgaacgccag	gcacccctgg
	63401	cccgcgcgg	ccagcaccgg	ggccggatgg	cagctgatcc	aggcggcgcc
	63451	ggtctcgacc	agccggttcc	agcgctggc	caccgtcacg	gcgtccagtc
	63501	cgaggacgga	gccgatccgg	gtccagctgg	cccgcggcgt	gatctgcagc
20	63551	gcgtgcacca	gggcctgatc	cacatggtcc	agggaccggg	gggtctgccc
	63601	ggaatcctgc	gccacgggcc	acctccttgc	gtgtttccgg	cggattcggg
	63651	ccgcccgtcg	gctcaacctt	cagcctggac	tcgggtacgg	ccggaccgta
	63701	ccaggcaacc	cccggagcaa	caggagtggg	attcggcatg	tcggcagtg
	63751	aacgcgccaa	ggaaatgcaa	cccagattgg	agcggctgcg	ccggtcgctg
25	63801	caccgggagc	cggagttggg	ccttgcgctg	ccgcgcaccc	aggagaaggt
	63851	gctcgcggcg	ctcgacgggc	tgccgttggg	gatcaccgcc	ggcaccggcc
	63901	tcacctcggt	gaccgccgtc	ctgcgcggcg	gccgccccgg	cggcgcagtg
	63951	ctgctgcgcg	gcgacatgga	cgcgtgccg	gtcaccgagg	agagcggtgt
	64001	gccgtacgcc	tcggagatcc	ccggccggat	gcacgcctgc	ggacacgacc
30	64051	tgcacaccgc	cggactcgtc	ggcgcgccca	ggctgttggc	cgaacgccgg
	64101	gaacgcctgc	acggcgacgt	ggtgttcacg	ttccagcccg	gcgaggaggg
	64151	ccacaacggg	gccggcgcg	tgatcgagga	gggcgtgctg	gacgcggcgg
	64201	gcaagccgct	cgacgcggcg	tacgccctgc	acgtggcgtc	caaccaggtg
	64251	cccggcgggc	tggtcatcac	ccgctccggc	accatcacct	cggcctccga
35	64301	cctgctgacg	gtgacgggtg	ggggcgaggg	cggccacggc	tcgacgcccc
	64351	ccacggccaa	ggacccgggt	ccggcgccct	gcgagatggt	caccgcgggtg
	64401	cagaactggg	tgacccgcgc	cttcgacatc	ttcgacccgg	tcgtgggtcac

64451 cgtcggcacc ttccacgccg ggacgaaggc gagtgtcatc cccgacaacc
 64501 gccgagttcc agggcacggc gcgcagttat tccgaatccg cccgggaccg
 64551 gctcgccgaa gggaaactgcc ccgtctggcg cgggccatcg cggccgcccc
 64601 cggcctggaa gcggaggcg actacctccg gcactatccg gtcaccgtca
 5 64651 acgacacgga cgagaccgcc ttcgcggcgg ccaccgcccg cgacctgttc
 64701 gcgcccggcg aggtgtggga gtcgccgata cccaccaacg gtcgagga
 64751 cttcgcgtag gtcctccggc gcgtccccgg cgcgatgctg ctgggtggcg
 64801 ccgccccga ggggagcgac tggcagcacg cgcgatgaa ccaactcgccg
 64851 cgcgccgtct tcgacgaccg ggtgctgtac cggcaggcgg cgctgctgac
 10 64901 cgaactcgcc gcgcgcggcg tcgcggccac cgcgcggccc gagcccgccc
 64951 tggcgggatg acaccgcccg cgcacgcgtg gggcccgcac gcgtgcggcg
 65001 tacgggccct ggccggtctc cttcccggct acttgcgcg ggggcaactcc
 65051 ttccacgcga agtggttagg ggtgttgatg ctgccgtcgg tggagtcag
 65101 gccatgtagc tgggtgcctt ggggtcagcc ccgctgccgt

15

SEQ ID NO. 2DNA sequence: Nys2

20 1 gaattcggat cgatgctgct cgtgggtgtg gccgcgggag gtgccgacgg
 51 tgccgtgctg gaatcggggc gtggtcgccg tcgcgatcgc ggtgggtgctg
 101 gtgctgctcg tgctgaccgt ggtgggtggcg acggcctact tgccgtgagc
 151 ccggtccttg cgcgcgaccc tgggcagcgg atgcacctct tcggtcgacg
 201 ccaagggggc gtggcgaggg tcgagatagg cgcggacgga gacgccgggc
 25 251 tgccccggcg tctgccaggc gcgcggaggg ccgctggaac gccacagacg
 301 gccagctcg gccgcggcaa ggggtggcgg ggttcgctcg cggcggtgga
 351 cgtcgatgac gaccatgccg ggttcgctga catgttctc gtggatatgc
 401 acagttctgt atgacctatt tcgccccgtg gcgtaaggag tagccggcag
 451 gtttcatccg aaggtggggc gcgggagcgg caggttgacc gtgaggtgcg
 30 501 tgccgtgttc ccggtcgcg cgggggggtg gggccggcag gtggaccgcc
 551 gtggtcagca gggcgggggc gctgtgccag cggccgggga agaccgtgac
 601 ccggcgaccg ccagcagggg ggcccgggac caggagtcgg gcggtgaagg
 651 cgccctgtgt cgggtcgatg acgatctcgg cctcgagaaa gtccagttcg
 701 cgctgggtca ggggatacca cgccctgaag acgctctcct tggcgtgaa
 35 751 gagcagccgg tcccagtgga cgtccggacg atgggcccgc agggccacga
 801 ggtgcggccg ttccgagggc agggcgatgg cgttcaggac gccggccggc
 851 agcggaccgt tcggttcggc gtcgatgctc accgcggccg acagctccgc

901 gggggagacg gcggcggcgc ggtagccggc gcagtgcgtc atgctgccga
951 cgatgcccgg cggccactgc gggcgccgc gccgattggg cagtatggcc
1001 cgggtccgggt ggccgagccg gcgcagggcc cggcgggcca gatggcggac
1051 ggtggtgaac tcgcgtgccc gggaactccac ggcccgggcg atgacctcgc
5 1101 gttccgagga gagcagccgg tcgccggggc gcggccggtc gtcgtacgcc
1151 gcctcgggtg cgaccgtggc gggcaggatc agttcgatca ccgcatacct
1201 ccggcgaacg gtaagaacag ggggttgctt cggggcacgg ttccgtcctt
1251 gacggggcgc cgtggggcgc cgggtcagcc ggcccgggcg cccgcgggtg
1301 gagtgtgggt gcgggctgcg tgcagccggg cgtacaggcc ctgcgcgcac
10 1351 aggagttgat cgtgggtgcc ctgctcgacg atgcggccgg cgtccatcac
1401 gacgatcagg tccgcgtcgc ggatcgtgga caggcgggtc gcgatcacga
1451 agctggtccg gcccgcgcgc agggagttca tggcgcgttg gatcaggacc
1501 tcgggtccggg tgtccacgga gctggtggcc tcgtccagca cgaggacggc
1551 cggctctggcg aggaaggccc gggccacggt cagcagctgc ttctcgccgg
15 1601 cgctgacggt gcccgaactc tcgtccagca cgggtgctga gccctgccc
1651 aggggtgcgga tgaagcggtc ggacacaggc gcgcggggcc cctcctcgat
1701 gtccgcacgg caggcaccgg gtgcgccgta cgcgatgttc tccgcgatgg
1751 tgccgccgaa cagccagggt tcctggagca ccagcccga gcgggaccgc
1801 aggtcgtcgc gggtcacgtc cgcgggtgctg gtgccgtcca ggaggatgcg
20 1851 gccggagtcc ggttcgtaga agcgcacag gaggttgccg aggggtggtt
1901 tgccggcgcc ggtggggccg acgatcgcca cgggtgctgc cggttccacg
1951 gtcagcgaga ggttctcgat gaggggcggtg tcgggggagt agcggaaagga
2001 cacgtcgggt aactcgacgc ggccctcggc gcggggcgggc gtgccggggc
2051 ggagcgggtc cggggcctgc tcggggcggt cgagcagggt gaagacgcgc
25 2101 tgggcggagg cgatgccgga ctggagccgg cccgccaccg aggcgatctc
2151 cacgatcggc tggctgaact ggcgggcgta gaggatgaac gcctgcacgt
2201 caccgagggt cagggtgccg tttatgacct tccaggcgcc gatgacggcc
2251 accagcacat agccgagggt ggcgacgaac atcatgaccg gttccatggc
2301 accggaggcg aactgcgcct tggccgcagc ccggtagacc gcgtcgttgc
30 2351 aggcgtcgaa gcgtcctcgc gccgccgcgc gccggtcgaa gcccttgatc
2401 agcgcgatgac cgggtgcacac ctcctccaca tgggcggtga gggtgccggt
2451 cgcggaccac tgcgcggcgt agtggggctg cgcgcgcttg ctgatccggg
2501 ccgcgatcaa cgcgcagacc ggcacgctga gcagcatcac cacggccagc
2551 gacggcgaga tcaccagcat cagcaccagc atcgtcaaca gcgagaagat
35 2601 cgagggtgatc agctcggcga ggggtctgctg gagggtctgt tggagggtgt
2651 cgatgtcgtt ggtggtgcgg ctgagcagct caccggccgg ctgccggctg
2701 aagtggcgca gcggcagccg ggtcagcttc tcccgggcgt cgcggcgacg

2751 ttctgtggatg gtgcgccaca ccgcggacgc caccagccgg ccctgcgcca
2801 gcatgaacag cgacgccacg acgtagagcg ccagcaggac cagcagcagc
2851 cggccgatcg cggcgaagtc gatcccgggc gccggggccc ggacgccggc
2901 gagcacgccg tcggcgatca gatcgggtgac ccggccgagc agcagcgggc
5 2951 cgaacgcgtt gagcacgac cgcgcgacgc ccatcgacac ggcgagtgcc
3001 acggagcggc ggtgcggacg cagcaggccg acgagccgac gtgccggccg
3051 gggcgcggtc cgctcctcct caaggtcgtc cggcgaggcc atggggcggt
3101 tcctcctcgg tcagctgcga gagcgcgac tcgcggtagg tcgggctggt
3151 gcgcagcagc acgtcgtggg tgccctgggc gaccacgcgc ccgcggctca
10 3201 gcaccacgat ccggtccgcg tcgcggccgg cggagatccg ctgcgccacg
3251 gtgatcaccg tggcgcccgc ggtgtacggc accagcgcgg tccgcagcgc
3301 cgcacgggtg gcctggtcga gcgccgagaa acagtcgtcg aagagataga
3351 tctccggccg gcgcagcaac gcccgggcga gggacaggcg ttggcgctgg
3401 ccgcgcgaga cattgccgcc gccctgggtg atctccgcgt cgaggccgtc
15 3451 cggcatccgc gccacgaagt cggccgcctg ggcgacccgc agcgccctcc
3501 acagtcctc gtcggtggcg tccggccgccc cgaagcgagc attgctcgcc
3551 acggtgccgg agaacaggta cggccgctgc ggcacgaacc cgacggcggc
3601 ggcgagcgtg gccgcggtca gtcgcggac gtcggtgccg ccgacccgca
3651 ccgcgcctc ggtggcgctc gccagccgca gcaccaagtt caacagggtc
20 3701 gtcttgccgc tgccggtgct gccgagcacg gcgatccgct cgcccggtc
3751 gacggtcagg tcgacgtccc gcagcacggg ctccctcggcg cccgggtagc
3801 ggtaccggc cgcgcacagt tcgatccggc cggcgggccc gcgcaccggc
3851 tcgggcgcgg ccggcgcgcc caccgtcgac ccggtgtcca ggacctccgc
3901 gatccggccg gcacagaccc gggcccgcgg caccgacagg aacacgaagg
25 3951 cgagcatcac gacggacatc aggatcagcg agagatagct caggaggggc
4001 ctgagcgagc cgatcggcat ccggcccgcg tcgatccggt gggagccggt
4051 ccacagcagg gctacggtga aaccgttcat cagcagcagc acgacccgca
4101 gcatcgtcgc gatcagccga cccaccgccc gcgacaccac gaggaacgcg
4151 tcgttggtct gcgcgaaccg cgcgcgctcg tggtcgtcgc ggacgaagga
30 4201 ccggaccacc cgaccccgg tgatcgctc gcgcagcagc cggcccagcc
4251 ggtccagggt cagctgcac cgcgcgtaca gggtgcccat ccgggcccagc
4301 agcaggccga agcagaccgc caccaccagc accagcgcca ccagcagcag
4351 tgccagcgga acgtcctggc gcagcgccag cagcacgctg cccaggcaca
4401 tcagcggcgc gcagacgacg atgccgaagc cggctcgggc gaggttctgc
35 4451 acctgctgca cgtcgttcac cgaccgggtc agcagggagg gggtgccgaa
4501 ccggccgatc tcggggcgcg agaagtccag gatgcggcgg aagagcgcgg
4551 accgcagatc gcggcccatc gccgtcgcgg tccggggcggc cagcgcggcc

4601 gcaccgagcg cggcggcgat ctgcaccagc gccaccacgc ccatcaccac
4651 acccagctcg gtgatgcgcc cgccgtcgcc gcgcaccacc ccctggtcga
4701 taagtgcggc gcccagcgtc ggcagcagca aagtgccag gatctggacg
4751 agttgaaggg cgacgagagc ggcgggtggcc caggcgtagg ggcgcagctg
5 4801 tgcccgcaga agtctcaaca gcacggagga acaccccggg tgacggcggc
4851 gtgcggccgc gcgggtggac gagtcggccg gcccgccgac ccggtcattg
4901 caccgcagtc cgctccgaaa ttctactagt gttgggggtg gcaacggact
4951 tttagcggcc cggtagctcg gaattcccta gaaagcccgg gatgcgttga
5001 cagcatcttc cgcggtcttc gagcgtgcgg gtgtctatcg tccggactgc
10 5051 gtattcgacg ccaggtcgtg gccgagctca agcgtttga cggctctctt
5101 cgtgttcgag aagggtttgc catgtccaaa cgagcgctga tcaccggaat
5151 caccggccag gacggctcct atctcgcgga gcacctgctg tcccagggtc
5201 accagggtgtg gggctctgat cgcggccagg ccaatccccg caagtccggg
5251 gtcagccgcc tcgcctccga actcgacttc atcgacgggg acctgatgga
15 5301 ccagggcagc ctggtctccg ccgtcgacac cgtgcagccc gacgaggtct
5351 acaacctcgg cgccatctcg ttcgtgccga tgtcctggca gcaggccgag
5401 ctggtcaccg aggtcaacgg catgggctgt ctgcgcatgc tggaagccat
5451 ccgcatggtc agcggactgt ccacctcccg cacggtcagc ccgcgcggcc
5501 agatccgctt ctaccaggcg tccagctcgg agatgttcgg caaggccgcc
20 5551 gagacgcccgc agcgcgagac caccctcttc caccgcgca gcccctacgg
5601 cgcggaag gcgtagggc actacatcac ccgcaactac cgcgagtcct
5651 tcggcatgta cgcggtctcc ggcatgtct tcaaccacga atccccgcgc
5701 cgcgccagg aattcgtcac ccgcaagatc agcctggcgg tcgcccgcac
5751 caagcaaggc ctccaggaca agctggcact cggcaacctc gacgcggtgc
25 5801 gcgactgggg ctatgccggc gactacgtcc gcgccatgca cctgatgctc
5851 cagcaggacg ccggcgacga ctacgtcatc ggcaccgggc agatgcactc
5901 ggtgcgcgac gcgggttcgga tcgcgttcga acacgtcggc ctgaactggg
5951 aggactacgt cgtcatcgac cccgacctgg tgcggcccgc cgaggtcgag
6001 gtgctgtgcg ccgacagcgc caaggcccag gaccgcctcg gctggaagcc
30 6051 ggacgtcgac ttccccaccc tcatgcgcat gatggtcgat tccgacctgg
6101 cgcaggtttc ccgcgaaaac caatacggcg acgtgctgct cgccgccaac
6151 tggtagcagt tctcaagctt tcgaaaacta gtgaattcct gccggaattc
6201 cgacgacact cgccacatgg attccccggg tgagtggcga atccagggtg
6251 cgaatccgaa cgtaccgccg aaacggcgtg gagaagtcgg acgccattca
35 6301 cgtgcggccg tcggctgtcg agatgggttg agttgagatg gacaacgaac
6351 aaaaactccg ggattacctc aagcttcgga cggccgacct tcgacgcacc
6401 cggcggcgcg tccacaagct ggagtcggcg gcccaggaac cggtggccat

6451 catcggcatg acctgtcgtt accccggcgg cgtccgcagc cccgaagacc
6501 tctggcgcgt ggtcgaggcc ggcgagcacg gcgtcaccac gttccccacc
6551 gaccgcggtt gggacctgga ggcgctggcc gccgcgccga ccgcctccgg
6601 cggattcctg caccgacgac ccgacttcga cgcggacttc ttcggcatct
5 6651 cggcgccgca ggcgggtcgcc atggaccgcg aacagcgcggt cgtcctggaa
6701 tccgcctggg aggcgttcga acgcgcgggc atcgaccgga cgtccgtgaa
6751 gggcagccgc accggagtct tcatcggcgc gatggcccag gactaccggg
6801 tcggccccgc cgacggcgcc gagggcttcc aactcaccgg caacaccggc
6851 agcgtgctgt ccggccgcat ctctacacc ttcggcacgg tcggccccgc
10 6901 cgtcaccgtc gacaccgcct gtcctcctc cctcgtcgcc gtccacctcg
6951 ccaccaggc gctgcgggcc ggcgagtga cctcggccct cgccggcgcc
7001 gtcaccatca tgtccggccc cggcaccttc atcgaaatgg gccgccaggg
7051 cgggctctcc gccgacggcc gctgcgcgtc cttcggcgac accgccgacg
7101 gcaccggctg ggcgaaggc gtcggcatcc tcgtcctgga acggtgtcc
15 7151 gacgccgtcc gcaacggcca cgagatcctc gccgtcgtcc gcggcaccgc
7201 cgtcaaccag gacggcgcc ccaacggcct gaccgcccc aacggccct
7251 cccagcagca ggtcatccag caggccctgg tcaacgccc actcggcgcc
7301 ggggacatcg acgtcgtcga ggcgcacggc accggcacca ccctcggcga
7351 ccccgctcag gcccaggccc tgctcgccac ctacgggcag aaccgcccgg
20 7401 cggaccggcc gctgtgtgtg ggctcggtea agtccaacct cagccacacc
7451 caggccgccc cggcgctcgc cggcgtgatc aagatggtea tggcgatgag
7501 gcacggcacc ctgccgcga ccctgcacgc cgaggagccc acccaccag
7551 tcgactggtc gcaggcgcc gtcgggtgc tgaccgacac caccgactgg
7601 cccgccaccg gggcgccgag ccgcgcggcc gtctcctcct tcggcatcag
25 7651 cggcaccaac gcccacacca tcacgagca gggcccgaa ccgcagccc
7701 aggacgccgc gaccgcgag gacgacgcc cgggcagcac gccggccacc
7751 gccccgtag tgcccgcggt cgtaccggtc ctgctctcgg gccgcacccc
7801 ggacgccctg cgcggccagg ccgcggccct gcgcgcggcc ctgcacaccg
7851 gccggcggcc cgacctgtc gacctcgac actccctcgc caccaccgc
30 7901 gccgggttcg agcaccgcgc cgtcctcctc gccaccgacc accccgcct
7951 gaccgacggc ctaccgccc tcgcgcacgc cgacgaccg gccgcgccc
8001 ccgcctggat caccggcacc acccggggcc agaccgggt cgcgctcctg
8051 ttcaccggcc agggcgccc acgcctcggc gcgggacggg aactcgccgc
8101 ccgtttcccg gcgttcgcca ccgcctcga cgcggcgctc gacgccttca
35 8151 cccgcacct cgaccgccc ctgcgcgagg tcctgtgggg caccgacgcc
8201 gccctgctcg accgcaccg atacgcccag ccggccctct tcgcgctcga
8251 agtggcgctc taccggctga tcgaatcgtt cggcgctcgc cccgaccacc

8301 tcgccggcca ctccgtcggc gagatcgtcg ccgcgcacct cgccgggggtc
8351 ctctccctgg ccgacgccgc caccctcgtc gccgcccgcg gtcgcctgat
8401 gcaggcgctg cccgacggcg gggcgatgat cgccgtccag gcgtcggaaag
8451 ccgacgtcgc cccgtgtctc gccgggcacg aggaccaggt cgcgatcgcc
5 8501 gccgtcaacg gccctccgc cgctcgctctg tccggcgccg aagccaccgt
8551 caccgcgctc gccgaacagc tcgccgccga cggccgcaag acccgccggc
8601 tgcgcgtctc gcacgccttc cactcgccgc tcatggagcc gatgctcgac
8651 gccttcgcg cgtcgctcga agacctcacg ctccagccgc cgctcctgcc
8701 ggtcgtctcc aacctgacgg gcaagcccgc caccgtcgcc caactcacct
10 8751 ccgccgacta ctgggtcgac cagtcgggc acgcccgtccg ctccgccgac
8801 ggcatcgact ggctcgcccc gcacgacacc accgccttcc tcgaactcgg
8851 ccccgcggc gtgctgtccg ccatggccca ggactgcctg gacgccggcg
8901 acgcagacgc cgtcaccctc cccgccctgc gcgcggggcg ccccgaggag
8951 cacaccctca ccaccgcct cgccgggtctg cagtcacacg gcgccaccct
15 9001 ggactggacc ggctgcttcg ccggcaccgg cgcccggcg accgacctgc
9051 cgacctacgc ctccagcgc cgccgtact ggccaaggc cctccagagc
9101 ggccacggcg acctgcgctc ggtcggcctc ggtgccgccc accaccgct
9151 gctctccgc gccgtctccc tcgccgacgc aggcggcacc ctgctcaccg
9201 gccgcctctc ccggcagacc caccctggc tcgccgacca caccgtccgc
20 9251 ggccaccacc tgctgcccgg tacgccttc ctogaactcg ccgtccgcgc
9301 cggcgacgag gtcggctgcg accgcgtcga ggaactcacc ctcgccgcac
9351 cgctcctgct gcccgaacag ggccggcgtc aggtccagtt gtggatcggc
9401 aaccccgacg tgtccggtcg ccgcaccgtc aacgtccacg cccgccccga
9451 caccggcgac gacacccct ggaccgcca cgccaccggc gtcctcacca
25 9501 ccgccgacgc ctcccgccag ctcccggctt cgctccgagca gggcggcacc
9551 cccctcgccg gcgaccccca ccccgccctc gacgcggccc agtggccccc
9601 ggccggcgcc gaaccgctgc cgctggacgg ccaactacgac cgcctcgccg
9651 acggcggtt cggctacggc ccggtcttcc agggcctgcg cgccgctgg
9701 cgccggggcg acgtcgtcta cgccgaggtc gagctgcccg aggcggggcg
30 9751 gtccgacgcc gaggcgttcg gcctccacc cgccctgctc gacgcggccc
9801 tgcacgccgc gcccttcacc ggctcgggc aacgcggccg gggcggcctg
9851 ccgttctcct gggagggcgt ctccctccac gccggcgggc ccaccacct
9901 ccggtccgc ctgaccccg tcgccgacga cgcgtcgcc ctgaccgtcg
9951 ccgacggcac cggcgcgccc gtgctgtccg tcgactcgct cgctcgtcgc
35 10001 agcgtggcga cccaacagct cgacacggcc gccgcccgtc cccgtgacgc
10051 cctcttccgc ctcgactgga ccccgctcca gccgaccgcc accgaccccg
10101 ggcccgtcgc cctcctcggc gccgaccct tcggcctgct caccacgcc

10151 ggattcgccg acgccccggc ataccggac ctgcccggc tcgcccggc
10201 ggacggcccc gtcccagca ccgtcgtgct gtccctcgcc ggcaccggg
10251 acgacggggc cgaccggcc cggtcgcac accgctgcgc cgcggaggcc
10301 ctgcccggc tacagacctg gctcgaccac catgagcgct tcgcccggc
5 10351 ccgcctggtc ttcgtgacct gtggtgcgac ggtcgggctg gatgttgctg
10401 ctgctgcggt gtggggtctg gtgcgttcgg cgcagtcgga gaatccggg
10451 tgttttgctc tggtcgatct ggatccggat ggtgcggtgg gtgcggctgc
10501 gctcgtcgct gcgttggtca gtggtgagcc gcagcttcgc gtgcgagggt
10551 atgtgttgcg ggtcgcgct ctggtgcggc ggccgctcac cgaggtcggt
10 10601 gcgggtgctg atggcaccgg ggatggcgct ggggatggct ctggtgtgtc
10651 gttctcgggt gaggtgctg tcttggtcac tgggtgtacg ggtggtctg
10701 gtgcggtgtt ggcgcgtcat ctggtggccg agtatggggt gcgggatctg
10751 ctgttggtca gtcgcagtg tgaacgtgcc gtgggtgctg gggagtgtgt
10801 ggcggagctt gcgggtgtg gtgcgagggt gcgggtggtt gcgtgtgatg
15 10851 tgaccgatcg tgcccggtg gtggagttgg ttggcgggca tgcggtgtcc
10901 gcgggtggtc atgcggtggt tgtgctggat gacggcatgg tgggtgcgtt
10951 gaccggggag cgggtgtccg cgtgctgcgc gccgaagggt gatgctgtct
11001 ggcattctaca tgaggcgacc cgcggcctg acctggacgc gttcgtcgtc
11051 ttctctctcc tcgcccgggt cttcggcagt cccggccagg ccaactacgc
20 11101 ggccgcgaac gccttctggt acgcgctgat gacgcggcgc cgggaggagg
11151 gactgcccgg cctgtcactc gcatggggac cgtgggagca gtcgggaggga
11201 atgacgggca ccctgacgga cgtcgacgcc gaacggctgg cccgctccgg
11251 tgtcccgcgc ctctccgtg cgcaggccct ggccctcttc gacgctgcgc
11301 tggccgggac cgacgccacc tgcgttcgg tccgctgga cctccccgtc
25 11351 ctccgcgcac ggggtgaagt gccgccgctg ctgaggctgt tgatccgctg
11401 ccgggcgcgc cgagccggc tcgccggctc cgccaccgcg ggcaacctcg
11451 cccagcgcct gcgcgcctg gacgaggac gccgcgacga gatggtcctg
11501 gacctcgtcc gcggtcaggt cgcctcgtc ctcgccacgc cgaccggtgg
11551 cgacgtcgac gccggcgtg ccttcgcga cctcggttc gactcgtgga
30 11601 ccgccgtcga actgcgcaac cgcctcaaca ccgtcaccgg cctgcgcctg
11651 cccgccacc tggctctcga ctaccgacc gtccggcacc tcgccacgta
11701 cgtcctggac gagttgttgg gcacggatgc cgagggtggc accgtgcagc
11751 cggccgcggt tgcggtggcg gacgatccga tcgtcatcgt gggcatggcc
11801 tgccgctacc ccggtggcgt cagctcccc gaggacctgt ggcgcgtgct
35 11851 caccgaaggc accgacggc tctcgggctt cccgaccaac cgtggttggg
11901 acgtcgaatc cctctatcac ccggaccctg accaccggg tacctctac
11951 acgcgctcgg gtgggttcct gcatgaggcg ggggagttcg atccggggtt

12001 cttcgggatg agtccgcggg aggcggttggc gaccgattcc cagcagcggg
12051 tgttggttga gtcgtcgtgg gaggcgatcg agcggggccg tattgatccg
12101 gtgagtttgc ggggtagtcg gacgggtgtg ttcgcggggg tgatgtacag
12151 cgattacagc gcgatgttg gcagtcagg gttcagagggt ttccagggca
5 12201 gtgggagttc gccgagtttg gcgtcgggtc gggtaggcta cacgttgggg
12251 ttggaaggcc cggcgggtgac ggtggatacg gcgtgttcgt cgtcgttggt
12301 ggcgatgcac tgggcgatgc aggcggttgc tagtggtgag tgtgggttgg
12351 cgttgcccg ggtgtgacg gtgatgtcga cgcctgcgggt gtttgtggac
12401 tttgctcggc agcggggttt gtcgccggat ggccgggtgca aggcggttgc
10 12451 ggatgcggcc gatggtgtgg gctggtccga gggcgtcggc gtgttggtcc
12501 tggagcggca gtcggacgcg gtgcgcaatg gtcacgagat tttggctgtg
12551 gtgccccgtt cggcgggtcaa ccaggatggt gcgtccaatg gtttgacggc
12601 gcccaatggt ccgtcgcagc agcgggtgat ccggcaggcg ttggccagtg
12651 gtggtctgac ggccgggtgac gtggatgtgg tggaggcgca tggtaggggt
15 12701 acgacgctcg gtgatccgat cgaggcgag gcgttgttgg cgacgtatgg
12751 gcgggatcgt gagcctgagc ggccggttgtt gttgggttcg gtgaagtcga
12801 atctggggca tacgcaggct gctgcgggtg tggcgggctg tatcaagatg
12851 gtgttgggca tgcggcatgg tgtggtgccg cggacgttgc atgtggatgc
12901 gccttcttcg catgtggact ggtccgaggg tgcggtgagg ctgctcagtg
20 12951 agcaggcggc ctggccggag acgggtcggg tgcggcgggc ggggtgtctcc
13001 tccttcggca tcagcgggtac caatgtgcat gtcacgtcg agcaggcgcc
13051 gggcgccaag gcgatcgccg cggccgggtgc ggcgcggcgc acgcggggtg
13101 ccgtgccgggt gctgctctcg gggcgtggcc ggagtgccct gcggggccag
13151 gccgcccgcg tgcctggaca cctccaggcc cgacccgacg ccgaactcgt
25 13201 cgatgtcgca ctgtcgttgg cgaccacccg gtcccgttc gagcagcggg
13251 ccgccgtcgt ggcgcaggac cgcgaccagc tgatcgccct gctgggggcg
13301 ctggccgccc accgccccga ccccgccgtc gtcgagggag aggcggccgg
13351 acgcggccgg accgcgggtc tgttcaccgg acagggcagc cagcgggccc
13401 ccatggggcg tgaactccac gaggtgcagc cggagtccgc cgcggcgctc
30 13451 gacgcggtgt gtgccgtttt cgaccgcgtg ttggaccggc cgtgcgcgga
13501 ggtggtgttc gccgaggacg gcagcgacga ggcgcactg ctggacgaga
13551 ccggttggac gcagccggct ctgttcgccg tcgagggtggc gctgttcggg
13601 ctggtggaga gttgggtgtt ccgtccggac ttcgtggccg gccattccat
13651 cggtagatc gcggcggcgc acgtcgccgg ggtgctgacg ttggaggacg
35 13701 cctgcggtct ggtggccgcg cgggcgacgc tgatgcaggc gctgccgacc
13751 ggcggcgga tgatcgcat ccaggccacc gaggacgaga tcgcggcga
13801 cctcgacgac acggtggcga tcgccgccgt caacgggccc cagtccgtgg

13851 tgatctccgg tgacgaggag gccgccgaaa cgatcgccgc cacgttcgcc
13901 gaacgcgggc gcaagaccaa gcggctgcgg gtgagccatg ccttcactc
13951 gccgcggatg gacgggatgc tggacgcctt ccggatcgtc gccgaggggc
14001 tgacctaccg ggcgccgcgc atcccgcctg tctccgacct caccggccgg
5 14051 cgcgccgacg atgcggaggt gtgcaccgcg gactactggg tgcggcacgt
14101 ccgagaggcc gtgcggttcg ccgactgcgt gcggacgctg cgcgacgcgc
14151 gggccaccac cttcctggaa ctgggctccg acggcctgct gaccgcgatg
14201 gccgaggaca ccctcggatg cgaccacgac gccgaactgg tgccgatgct
14251 gcgcgccggg cgcgcggagg aactggccgc ggccaccgcc ctggcccgcg
10 14301 tccaggtgcg cggcgtggac gtggactggg cggcgtaact gcccggcacc
14351 ggcgcccacg gcaccgacct gccgacctac gccttcacgc acgcgtacta
14401 ctggccgcag ctggcgacct cggccgcgcg cctcgccgcc gccgatcccg
14451 ccgaccagca gctgtgggac gctgtggagc gcggcgacgc ccgcgaactc
14501 gccgacatcc tcggcctggg cgaacaggac ctacgcgcgc tggactccct
15 14551 gctgcccgcc ctacactcgt ggcgccgcgc caaccaggag aagcacctcc
14601 tggacaccct gcgctaccgc gtggagtggc cacgactgag caagccgacc
14651 gccccggtcc tcgacggcac ctggctgctg gtcgcctccg acgccaccgc
14701 ggccgaccag ccagccctcc tcgacggcct ggccgacgcc ctcggtctgc
14751 acggcgcgcg ggtgcgtcgc ctgcttctgg acgactcctg cgcggaccgc
20 14801 gcggtgctcg ccgaacgact ggcgccgacc gccgacgtgg acgccgcgac
14851 ccaggtgctg tccgtgctgc cgctcgacga gcgggacgcc gacgactgcc
14901 cgccgctcac ccgcggactg gcgctgaccg tcgcgctcgt ccaggccctc
14951 gccgacaccg gcgccaggg ccggtgtggt accgccacct gcggcgccgt
15001 ctccaccaac ccgcgcgacc cggtcaccca ccccgctccg gccgctgcct
25 15051 ggggcctggg ccggggcgtc gccctggagc acccacggct gtggggcggc
15101 ctggtcgacc tgccgcaggt ctcgacgag gcggccggac agcggtctgc
15151 cgggatcctc gccgtcaagg acgcaccgga gcggcaggac cagggtggcg
15201 tgcgggccac cggagtctcc ggccgcgggc tcgtccgcca caccgtcgaa
15251 gcgctgcccc cggcccgagg gttcacccgc accggcactg tctgatcac
30 15301 tgggtggcac ggtggcctgg gcgccgaggt cgcgcggtgg ctggcccgcg
15351 ccggcgccca gcacctcgtc ctgaccagcc gcccgggccc ggacgcgccg
15401 ggcgccgccg aactccgggc cgaactggag ggctacgggc cgtcgggtgc
15451 cgtcgtcgcc tgcgacgtc ccgaccggga cgcgctcgcc gccgtctca
15501 ccgactgcc cgaggaaact ccgctgaccg gtgtcgtgca caccgcaggc
35 15551 gtcggccact acggcccgtt ggacaccctg agcaccgccg agttcgccgg
15601 cctcacgcgc gccaaagctc ccggcgccgc ccacctcgac gccctgctcg
15651 ccgaccgcga actggacttc ttcgtctctt tcggctccat cgcgggtgct

15701 tggggcagtg gcaaccagag cgcctacggc gccgccaacg cctacctcga
15751 cgcgctcgcc ctgcaccgcc gcgcccgcgg cctcgccgcg acctccgtcg
15801 cctggggccc gtgggcccag gccggcatgg ccgcccacga tgccgtttcc
15851 gagaccctgc gccgccaggg cctcggcctg ctcgaccggg ccccgcccat
5 15901 gaccgagctg cgccgcgcgg tcgtccggca ggacgtcacc gtcaccgtcg
15951 ccgacgtgga ctggcagcgc tacgcaccgc tggtcacctc cgccccggcc
16001 agcgccctga tcgccggcct gcccgaggtc cgcgccctcg ccgcccacga
16051 gcgcaccgag caggacgcca ccggcgccctc cgaggtcgtc acccgcgctc
16101 gcgccctggc cgaaccgcag caactgcgcc tgctgaccga cctcgctccg
10 16151 accgagtccg ccaccgtcct cgccacagc tcgcccgcag ccgtgcccga
16201 gggccgcgcc ttccgcgacg tcggcttcga ctcgctgacc gcggtcgagc
16251 tccgcaagcg cctgggcgcc gcgaccgggc tgtccctgcc cagcaccatg
16301 gtcttcgact acccgacacc gctggaactc gccagttacc tgccggcgga
16351 gatcctcggc gcggtgctgg aagtcgccgg cccggtcgcc accggcgggc
15 16401 ccgacgacga gccgatcgcc atcatcgga tggcctgccg cttccccggc
16451 ggcgtcagct ccccggaaca gctgtgggac ctggtcgcc cccgaccga
16501 cgcgatcagc gagttccccg tcaaccgcgg ctggcagacc gggcacctct
16551 tcgaccggga ccccgaccgg cccggcacca cctactccac ccaggcgggc
16601 ttctccacg aggcgcagca gtcgacccc accttcttcg gcatctcgcc
20 16651 ccgcgaggcg ctggtcatgg acccgagca gcggctcctg ctggagacca
16701 cctgggagtc cttcgagcgc gccgggatcc gcccgaaac cctccgatcc
16751 accctgaccg gcaccttcgt cggctccagc taccaggagt acggcctggg
16801 cgcgggcgac ggcaccgagg gccacatggt caccggcagc agccccagt
16851 tgctctccgg ccgactgctg tacgtcttcg gtctggaagg cccggcggtc
25 16901 accgtcgaca ccgcctgctc gtctcgctc gtggcgctgc acctggcctg
16951 ccagtcgctg cgcaacggcg agagcaacct ggccgtcgcc ggcggcgcca
17001 cgatcatgac gacgccaac ccgttcacg cgttcagccg gcagcgcgcc
17051 ctcgccaagg acggccgctg caaggcgctt tccgacgacg cggacggcat
17101 gacgctcgcc gaggcgctg gcgtcgctc cgtcgagcgg ctctccgacg
30 17151 cgcagcgcaa cggccaccgg gtcctggcgg tcctccgcgg ctccgccatc
17201 aaccaggacg gcgcctcaa cggcctgacc gcgccaacg gccgtccca
17251 gcagagggtc atccgccagg cctcgccaa cggccgctc gcggcgggg
17301 acatcgacgc cctggaggcg cacggcacgg gcacaccgct cggcgacccc
17351 atcgaggccc aggcactgtt cgccacctac ggccgcgacc gcgacccga
35 17401 gagcgcgctg ctgctcggtc cgggtgaagtc caacatcggc cacaccagt
17451 ccgcccggg catcgccagc gtcacaaaga tggcatggc gctgcgccac
17501 tccgaactgc cgccgaccct gcacggcgac gcggcgctc cgcacgtgga

17551 ctggctcggcc gggacgggtcc ggctgctgac ccaggcgcgcc gcctggccgg
17601 agaccgggtcg cccgcgccgg gccgcggtgt cctcggttcgg catcagcggt
17651 accaacgccc atgtcctgct ggagcaggcg cccgtcgcg acaccccgcc
17701 cgaggagcgg cccgcctggg cgcgggtccc gatcgccgcc ggcgtcgtec
5 17751 cgtgggtggg caccgccccg agcgccgccg ccctgcgcgg ccaggccgag
17801 cgcctcctcg cgcacgccga aaccgtcgga accgccctgc cggccgcgg
17851 accgctcgac atcggcctgt cgtgggtctc cgcgcgcgcc cgtttcgagc
17901 accgtgccgt cgtcgteccg cccgcgggca ccgaccgct ggcgcgctg
17951 cgcgcgctcg cgacggacgg gccctcgccc gtggtcgccc gtggcgtggc
10 18001 ggacgtcgag ggtcggacgg tgttcgtgtt ccccggtcag ggttcgcagt
18051 ggggtgggat ggggtcccaa ctcttgatg agtcggcggt gttcgcggag
18101 cggattgccg agtgtgcggc ggactcgcc gagttcaccc actggtcgct
18151 ggtcgatgtg ctgcggggtg tgggtgggtgc gccgtcggtg gagcgggtcg
18201 atgtggtgca gccgcgctcg ttcgcggtga tgggtgctgt ggctgcgttg
15 18251 tggcgttccc gtggtgtgtt gccggatgcg gtggtggggc attcgaggg
18301 tgagatcgct gctgcggtgg tgcggtgct gctgtcggtg cgggacgggg
18351 cgcgggtggg ggcgctcgcg agtcaggcca ttggtcgctg gttggcgggg
18401 cggggaggga tgatgtccgt cgcgctgtcg gtggacgtgc tcgaaccgcg
18451 gttggtcgag ttcgaggggc ggtgtcggt ggccgccgct aacggcccgc
20 18501 gctccgtcgt ggtcgccggc gagcccgagg cgctggacgc gctgcacgcc
18551 cggctgaccg ccgacgacat ccggggccgc cggatcgcg tggactacgc
18601 ctgcactcg caccaggctc aggaacctga cgaggaaactg ctggagggtgc
18651 tggcggagct ggcgcgcgc acgtcggagg tgccgttctt ctgcaccgtg
18701 accggcgact ggctggacac cgcgcggatg gacgcggct actggttccg
25 18751 caacctgcgc ggacgggtgc ggttcgcgga cgcggtggcg gacctgctgg
18801 cggcggagta ccgcgcatte gtcgaggtca gctcgacccc ggtgctgtcg
18851 atggcgggtgc aggaggcgat cgacgaggcc ggcgtgccgg ccgtcgccgc
18901 cggcaccctg cgcgcgacc agggcggcac cgaccgcttc ctgctgtcgg
18951 ccgccgaggt ctctgtgcgc ggtgtggacg tggactgggc ggggtgttc
30 19001 gaggggaccg gtgcgtcccg gatcgacctg cccacctacg ccttccaaca
19051 cgaacacctg tgggccgtcc cgcgcgccc ggaggccgct gccgcgccg
19101 acccggacga cgcggccttc tggaccgcgg tcgaggacgg tgacgtctcc
19151 gcgctgaccg ccgcgctcgg caccgacgag gactccgtcg ccgccgtgct
19201 gcccgccctg acctcctggc gccgggccc cgcgaccgc tccaccgtgg
35 19251 acgcctggcg ctaccgcgtc gcctggaaac ccctcggcgg caccctgccg
19301 caccgtccc tgaccggcac ctggctgctg gtcaccgcc acggcatcga
19351 cgacaccgat gtggcagggg cgttgagac ctacggcgcc gaggtgcgcc

	19401	ggctggtcct	ggacgaggag	tgcgtcgacc	gcgccgtcct	gcgggagcgg
	19451	ctggccggcg	cggaggacgt	gaccggcatc	gtctccgtcc	tcgccgcgc
	19501	cgagcggacc	gacgcggtac	cgggcacctc	cctggtgctc	ggcaccgccc
	19551	tgaccgtggc	actgatccag	gccctcggcg	acgccgaaat	cgacgtctcc
5	19601	gtatgggctg	tgaccgcggg	cgcggtctcc	accggccggg	ccgacgagct
	19651	gaccgcgccc	gtccaggcac	aggtcaccgg	catcggtcgg	accgcggcgc
	19701	tggagcacc	gcagcgtgg	ggcggcacc	tcgacctgcc	cgccgcctc
	19751	gacgcccggg	cgcccagcg	gctcgccgcc	gtgtgtctcg	gcgccctcgg
	19801	cagcgacgac	cagctggcca	tccggccctc	cggggtcttc	accgcgcga
10	19851	tcgtgcgggc	cgaggccacc	gccgggcggc	ccgccggcac	ctggacgccg
	19901	cgcggcacca	cactggctac	cggcggtccc	ggcaccctcg	ccccgcacct
	19951	cgcccgtgg	ctggcccaac	gcggcgccga	gcacctggtc	ctgatcagcc
	20001	ggcgcggcac	ggccgccccg	ggcgccgccg	aactcgtcgc	ggaactggcc
	20051	gagtcgggca	ccgaggcgac	cgtcgccgcc	tgcgacatca	ccgaccgcga
15	20101	cgcggtcgcc	gcgctgctgg	ccgacctcaa	ggccgacggg	cgcaccgtcc
	20151	gcaccgtcgt	gcacaccgcc	gccaccatcg	agctgcacac	cctggacgcc
	20201	accaccctgg	cggacttcga	ccgggtgctg	cacgccaaag	tcaccggcgc
	20251	ccaggtcctc	gccgaactgc	tcgacgacga	agagctggac	gacttcgtcc
	20301	tgtactcctc	caccgccggc	atgtggggca	gcggcgccca	cgccgcctac
20	20351	gtcgccggca	acgcctacct	cgccgcgctc	gccgagcacc	gccgggccaa
	20401	cggactgccc	gccctgtcgc	tgtcctgggg	catctggggc	gacgacctca
	20451	aactgggccc	ggtcgatccc	cagatgatcc	ggcgagcggg	cctggagtcc
	20501	atggaccggc	agctggccct	gagcggcctg	cagcgggcgc	tggacgacaa
	20551	cgagaacgtg	ctcgcggtcg	ccgacgtgga	ctgggagacc	taccaccccg
25	20601	tctacacctc	cgcccgaccc	accccgctct	tcgacgaggt	gccggaggtc
	20651	cgccggctca	ccgcggccgc	cgagcagagc	gccgggaccg	tcgccgaggg
	20701	cgagtctgcc	gccgcgctgc	gcgccctgtc	cgacgcccag	cagcagcgca
	20751	ccctgctgga	gaccgtccgc	accgaggcgg	cgtccgtcct	cgggctgtcc
	20801	tccgccgagg	acctcaccga	ccagcgggcc	ttccgcgacg	tcggcttcga
30	20851	ctcgctgacc	gccgtcggcc	tgcgcaaccg	gtcgcctccc	gtcaccggcc
	20901	tgacgtgccc	ctcgacgatg	gtcttcgact	accccaaccc	ggccgcgctc
	20951	gccgcctatc	tgcacggcga	gctggccggc	gcccggctccg	ccgccgcggg
	21001	cgccgccgcc	gtcccgaccg	gcgccccga	cgccgacgac	ccgatcgcca
	21051	tcgtcggcat	gagctgccgc	taccccgggc	gggtcggctc	cgccgaggac
35	21101	ctgtggcgga	tcgccctgga	cgaggctcga	gcgatctccg	gcttccccgc
	21151	cgaccgcggc	tgggacgccg	agggcctcta	cgacccgga	cccgaccggc
	21201	ccggccgcac	ctactccgtc	cagggcggat	tcctgcgcga	cgtcgccgag

21251 ttcgacccgg gcttcttcgg gatctcgccg cgcgaggcgc tgcgatgga
 21301 cccgcagcag cggctcctgc tggagaccgc ctgggaggcg ttcgagcacg
 21351 ccggcatcga cccggtcggc cagcgcggca gccgcaccgg caccttcgtc
 21401 ggcgccagct accaggacta cgcctccggc gtgcccaaca gcgagggctc
 5 21451 cgaaggccac atgatcaccg gcacgtcttc cagtgtgctg tccggccggg
 21501 tgtcctacct cttcggcttc gagggcccg ccgtcacgct cgacaccgcc
 21551 tgctcctcct ccctggtcgc catgcacctg gcctgccagt ccctgcgcaa
 21601 cggggagagc tcgctggccc tggccggcgg cgtcagcatc atgtccaccc
 21651 cgatgtcgtt cgtcggcttc agccggcagc gcgccctcgc cgaggacggc
 10 21701 cgctgcaagg cgtacgcgga cggcgccgac gggatgacct tcgccgaggg
 21751 cgtcggcctg gtgtgtgtgg agcggctgtc cgacgcccgc gccaacgggc
 21801 accaggtgct cgcctgatac cgcggctccg ccgtcaacca ggacggcgcc
 21851 tccaacggcc tgaccgcacc caacggcccg tccagcagc gcgtcatccg
 21901 ccaggcgtg gccaaactcc cgggtggcgc cggcgacatc gacgtcctgg
 15 21951 agggccacgg caccggtacc gccctcggcg accccatcga ggcgcaggcc
 22001 ctgctcgcca cctacggcca ggaccgcgc cccgaacggc cgctgtgtgt
 22051 cggctcggtg aagtccaaca tcggccacac ccagatggca tccggcgtcg
 22101 ccagcgtcat caagctcgtc cgcgccctcc aggaaggcgt ggtgcccaag
 22151 tccctgcaca tcgaccggcc ctccaccac gtcgactggt cctcgggcgc
 20 22201 categggctg ctcaccgaac gcaccccggt gcccgagacc ggccggccgc
 22251 gccgcgccgc cgtctcctcc ttcggcatca gcggcaccaa cgtccacacc
 22301 atcctcgaa c agggcccgcc ggacgaggcg cccacgccc cgcaccgccc
 22351 gcgggacggc ctggtgccgg tctgtctctc cggccgcggc gaggccgcgc
 22401 tgcgcgccc a ggccgcccgc ctgctcgctt tcgtcgagga gcggcccag
 25 22451 gccacctca ccgacctgc cactccctc gccacctgc gcgccgcgt
 22501 ggaacgccgc gccgcggtca tcgcccgca ccgcgacacc ctgaccgcgc
 22551 gcctgcgcgc cctgtccgac ggccggcccg accccggcct ggtccagggc
 22601 accgcgggac gcggccggac cgccttctct ttcaccggac agggcagcca
 22651 gcgccccggc atgggcccgc aactccacga ccgctaccgc gtgttcgccg
 30 22701 acgcgctgga cgaggtgctg gcccggtcgc acgacggacc ggaccggccg
 22751 ctgcgcgagg tgctgttcgc cgcgcccgac tccgccgagg ccgcgctcct
 22801 ggaccggacc ggctacgccc agcccgcgct gttcgccgctc gaggctcgcgc
 22851 tgttccgcct gctgacgtcc tggggcctga ccccggaact cctggccggc
 22901 cactccgtcg gcgaactgc cgcgcgcac gtgcgcggcg tgctgtcgt
 35 22951 ggacgacgcc tgactctgg tcgccgccc cggccggctc atgcaggcgc
 23001 tgcccagagg cggcgcgatg gtcgccctgg aggcgcgga ggacgaggtc
 23051 ctgccgctcc tggaggggct caccgaccgg gtgtccgtcg ccgccgtcaa

23101 cgggccgcgg tccgtggtcg tcgccggcgt cgaggaggac gtgctectcc
23151 tcgccgacct cttcgccgcc gacgggcgcc ggaccaagcg gctgcgggtg
23201 agccatgcct tccactcgcc gctgatggac gccatgctcg acgacttcgc
23251 cgccgtcgcc cgcgggctga cctaccaccc gccgacgac ccgttcgtgt
5 23301 cgaacgtcag cggcggcctg gccaccgccg aacagggtccg cacgcccgcac
23351 tactgggtcg ggcacgtccg cgcgcgggtc cgcttcgccg acggcatcga
23401 ctggctcgcc acccagggcg acgtccacac cttcctggag ctcgggcccg
23451 acggcgtgct cagcgccatg gcccgggaga gcctcaccga cccgtcccgc
23501 acggcactgc tgccgacct gcgcggcgac cggcccagg aacctgccct
10 23551 ggtcaccgcg gtcgccgcgg ccacgcgca cggcgcccgc gtcgactgga
23601 gcgggtactt cgcgaccac ggcgcgcgcc ggaccacgt gccgacctac
23651 gcgttccaac gcgagcggta ctggcccgc accaccgcc ccacgagcgc
23701 ccacacgccc ggatccgccc tcgacgccga gttctgggcc gccgtcgagc
23751 gggacgacgt cgcgccctc gcgcctccc tggacctgga cgacgccacc
15 23801 gtcaccgcga tggccccgc gtcaccgcc tggcgccgc gcccggcga
23851 gcagaccgaa ctggactcct ggcgctacc cgtcacctgg aagccgcgcg
23901 gcggcgccac cgcaccgcc gccctcacc gccgtggct cgtgctcgtc
23951 ccgcacgacc accaggacc tcaggacgac gcgaccgcg cctgggcagc
24001 cgacgtcgag accgccctgg gcaccaccac cgtccggctg acggtcacca
20 24051 ccaccgaccg cgcgcgctg gccgcccgga tcaccgaagc cgccggcgac
24101 cagggcccgt tcagcgggtg gctgtccctg ctgccgctcg ccaccggcga
24151 cgccggccac cccggtgcgc ccgcgcctt caccctcacc accaccgcg
24201 tccaggccct cggcgacgcc ggcacgcgac cgccgctgtg gaacgtcacc
24251 cgcggggcgc tggcgcgtcg ccgcgccgaa caggtcaccg cgcccgaaca
25 24301 ggccgcccgc tggggcctgg gccgcgccgt cgccctggaa ctgccggccc
24351 ggttcggcgg caccctcgac ctgcccgcga ccctggacgg ccaggccgcc
24401 cgcgggttgc gcgcggtgct cgcggctacc gacggcgagg acgcggtggc
24451 cctgcggccc tccggcgtct tcctccgcg cctggcccac gcccggccg
24501 gccccgacac cgccgcacc gccttcgacc cggcgggccg caccgtcctg
30 24551 atcaccggcg gcaccggcg catcgggcg cagtcgccc gccgcctggc
24601 ccgcgacggc gccaccacc tgcgtgtcac cagccgcgc ggccggccg
24651 cccccggcgc cgacgcgctc cgcgccgaac tggaggaaact gggcgcccg
24701 gtcaccctcg ccgcctgcga cgccgcgac cgcgacgcgc tggcgcgct
24751 cctcgccgaa ctgcccgac agccccgct gtgcgcggtg ttccacaccg
35 24801 ccggcgctgt cgaggaccac gtcgtggac cgctcacacc ggagaacttc
24851 gccgcggtgc tgccgcgcaa gaccgtcgcc gccaccacc tgcacgagct
24901 gaccgcccgc ctggacctcg ccgcttctgt gctgttctcc tccacggccg

24951 gcgtcctcgg cgccgccgga cagggcaact acgccgcggc caacgcccac
25001 ctggacgccc tcgccgaaca ccgccgtcc caccggcctga ccgcgctgtc
25051 cgtcgcctgg ggcccggtgg ccggctccgg catggtcgcc gacgccgccc
25101 aactcaccga ccgggtacgg cgccggcggt tcgaaccgct ccccccgaa
5 25151 ccggccgtgc gcgccctgct gcgcgccatc gagaacgacg acaccaccgt
25201 cgcgctcgcc gacatcgact gggagcgctt ccagcgcgcc ttcgccgagg
25251 tccgcccgct gccgttcgtc gccgacctcc ccgagaccgg ccggggccacc
25301 cccgcgaccg ccaccggcgc gccaccggc ctgcggcagc aactcgccga
25351 actgccggag caccgagcgc cggcagcggc cctggacctg ctgcgtaccc
10 25401 aggtcgccgc cgtcctcggc caccgccgacc cgccgaccgt cgaggacgac
25451 caccgcttcc gcgacctggg cttcgactcg ctgacctacc tggaaactgcg
25501 caacgcccctc aacgccgcca ccggcctgag cctgcccggc accctgggtct
25551 acgacctgcc caccgcgcgc gagatggcgg acttcctgct cgccgaactc
25601 ctcggcaccc tgcccaccga caccgcgcgc accgtcgcca gcacggcctc
15 25651 ccccaagctc tcagcttcgt tcgagcaggc cggtagcccc ttcgacgacc
25701 cgatcgccgt catcgccatc ggctgccgct tccccggcgg cgtcaccacc
25751 ccggaggagc tctggcagct cctcgacgag ggccgcgacg gcatcagccg
25801 cttccccgac gaccgcggct gggacctcgc cgcgctgggc gccggcgctt
25851 ccgacaccct ggaggggcgc ttctgaccg gcgtcgccga cttcgacgcc
20 25901 cggttcttcg gcatctcgcc ccgcgaggcg ctggccatgg acccccagca
25951 gcggctgctg ctggagacca cctgggaggc gctggagcgg gccggcatcg
26001 acccgaccac gctgcgcggc tccaccaccg gcgtcttcgt cggcaccaac
26051 ggccaggact acccgacgct gttgcgcgc tccgcctcgg acgtggccgg
26101 ctacgtcgcc accggcaaca ccgccagcgt gatgtccggc cgctgtcctt
25 26151 acgcgctcgg cctcgaaggc ccggccgtca ccatcgacac cgctgtctcc
26201 tcctcgctcg tcgccctgca ctgggcccgc cgggcgctgc gcgccggcga
26251 gtgcgacctc gtggtggcgg gcggcgcttc ggatcatggc agcccgact
26301 ccttcgtcga gttctccacc cagggcgggc tggcaccga cgggcgctgc
26351 aaggcgcttct ccgacgccgc cgacggcacc gcctggctcc aaggcgctcg
30 26401 catcctcgtc ctggaacgcc tctccgccgc ccgccgcaac ggccaccagg
26451 tcctcgccct gatccgcggc accgcgctca accaggacgg cgctccaac
26501 ggctgaccg cgcacaacgg cctctcccag cagcgctca tcgcccaggc
26551 actcgccgac gcccgctgc gcccgccga catcgacgcg atcgaggcgc
26601 acggcaccgg caccaccctc ggcgaccga tcgaggcccg cgccctgatc
35 26651 accgcctacg gccgggaccg ggacgcgaa cggccgctgc tgcggggcgc
26701 cgtcaagtcc aacatcgggc acaccaggc cgccgcgggt gccgcggcg
26751 tcatcaagat gctgatggcg atgcgccacg gcaccctgcc caggacgctg

26801 cacgtgggca ccccgctccag ccacgtcgac tggagcggcg gcaccgtcgc
 26851 gtcctctgac gacgcgcggc cctggccacg gaccgggcag ccgcggcgcg
 26901 ccggcgtctc cgccttcggc gtcagcggca ccaacgcca cgctcgtcgc
 26951 gagcaggccc cggaaaccga agcccccgcc gccccggccg ccgagccggc
 5 27001 gccggaggcc acgcccaccg tcgtcccctg ggctcgtctcc ggacgcagcc
 27051 gggaaagcgt ccaggcgcag ctggaccggc tcaccgcgca caccgccgcc
 27101 caccgcgcgc gctcggcggc ggacgtcggc cgctcgtcgg ccaccgaccg
 27151 cacgctcttc ccgcaccgcg ccgtgctgct cgccggcccg gacgggggtgc
 27201 gcgagggcgc ccgcgcgcgc gcgccccgca cccccggccg caccgcgttc
 10 27251 ctgttctccg gacagggcgc ccagcacgcc ctgatgggccc acgacctgta
 27301 ccagcgcttc ccggtctacg ccgacgcact ggacaccgtc ctgcgccagt
 27351 tcgacaccgt gctggacgtc ccgtgcgcgc ccgcgctggt cgccgcgcgc
 27401 ggcacccccg aggccgcgct cctggaccag accggcttca cccagccgcg
 27451 gctgttcgcc gtcgaggtcg cactgttccg gctcgccgag tcctggcggc
 15 27501 tgacgccgga cttcgtcgcc ggccactcca tcggcgagat c

SEO ID NO: 3

NysD2 (incomplete - probably C-terminally truncated).

20

1 msftypvsmp wlqgreldyv teavgggwis sqgpyvrrfe eafaayndvp
 51 fgvacssgtt altlallralg vpggdevivp eftmiasawa vtytgatpvf
 101 vdcgddlnid vsrieekitp rtkvimpvhi ygrqcdmdav lnlayeynlr
 151 vvedsaeahg vrprgdiacf slfankiisa geggvclthd phlaeqmahl
 25 201 ramaftkdhs flhkklaynf rmtnmqaava laqteqltdi lalrrdiekr
 251 ydealrdipg itlmpprdv l wmydlraerr delcaylage gietrvffkp
 301 msrqpgyfsa dwpalnaarl sadgfylpth tgltaqeqef itgri

SEO ID NO: 4

30

NysD1

1 vtlpsgntrl gwrrrrmhsp gdragrvrga rarrpatfrg vlsmganrrp
 51 ilfvsysaesg llnpillvag elsrrdvadl.wfatdekard evaavvdgsp
 35 101 vrfaslgdtv sqmsavtwdd atyaevtqrs rfkahaavir hsfapesrma
 151 kyrrleeive evepalmvie smcqfgyela itkgipfvlv vpfvpsnvlv
 201 shvpfaksyt psqfpvphsg lpaamslaqr ienqlfrlrt lgmfltsdvr

251 kvveednrvr telgiapqar qmmaridhae qvlcysvrel dypfpmhpk1
 301 rlvgtmvpp1 pqapdddgl5 dwlsaqsxv ymgfgtitrl treqvaslve
 351 varrlldgrgh qvlwklprgq qellppaael pdnlriegwv psqldvlahp
 401 nvkaffthag gngyheglyf gkplvvrplw vdcddqairg qdfgvs1tld
 5 451 rpetvdtedv ldkitruldq psfteraehf agllrdaggr aaaad1llgl
 501 palatd

SEO ID NO: 5

10 NysA

1 mtigadedpv vvvgmacryp ggvagpedlw elvrtgrdat tafpddrgwd
 51 laalagdpgp rsatreggfl tgaadfdaa1 fgmspreavs tdpqqrlvle
 101 tawealerag idphslrgsr tgvfvgasgq dyaavthasp ddldghaltg
 15 151 lapgvasgrl ayvlglegpa vtvdttssss lvalhwavra lragecstal
 201 aggvvmstp aafvghtrqg glapdgrckp fsddadgtaw aegvgivvle
 251 hlstaraagn pvlavlr1gsa vngdgasdgl tapsgpqer viraaladar
 301 lapadidlve ahgtgtrlgd pvearallaa ygqdrdpdrp lrlgslkstl
 351 ghaqaaagig gviktvl1tlr hglmprirhl atptrqv1dws qgavaplt1dh
 20 401 t1p1wppadrpr ragvssfgis gtnahvilee appadvp1vtr pgtl1rpstvp
 451 w1p1vsaatpea l1daqlarlra h1lrthsdldp l1dvgy1slatg raalrh1rav1
 501 l1ppadgtaad avehargaah qrrtav1lfsg qgsqrpgm1gr elaarfpvfa
 551 dalddalral drhldg1pvre vmwgtdaall d1rtgwtqpal favevalhrl
 601 vaslgvtpdf vgg1hsvgeia aahvagv1sl edacrlvaar atlmqalpag
 25 651 gamaaleate devapllgah lalaavngpt avvvagaeda vrqltarfad
 701 rgrrtsrlav shafhsplme pml1dafrdv srltfhqpsi plvsnl1tgel
 751 agseitsaey wvrhvr1d1tvr fadgitalak agadvlielg pggvlsamar
 801 dtl1gpdsttd vvpalskgrp eetafagalg rlhtlgvpvd wpafyagtga
 851 rrvelptyaf qhvrhwptpp r1pngagpgal ghpllgsave ladgggtvcs
 30 901 galslrthpw ladhtvagrv vlpatallel avragdeagc dvlhelhltt
 951 ppalpddaal hvqvhvgpad ttgrravtvh trpdhhpagd wtrcatgtlg
 1001 stppsaa1eaa tggtpaawpp adaep1dlad hyerladrgf dygptfrglr
 1051 aawrrgaeif advecppgta ddapdhglhp all1daarhaa mavdgtvpva
 1101 whgvrlhav1g atalrvrirp tttgtl1lta vdvhgapvvt vealtarplt
 35 1151 deeraaprt1p rqargetpad arparpaaar pgpageplpd ttgshptagh
 1201 laalppaare rqlldlvrtq aaavlg1hpgp eavgtrsvfk elgfdslagv
 1251 eladr1ltart glrlpatlvf nfptperaah rlgellaata pldpgaygee

1301 ltrfeaivtn lpqdgpperra vadrldaivs alrqnspacev pssdedidtv
 1351 svdrllidiid eefett

SEQ ID NO: 6

5

NysB

1 mgepqggqpd qqekivdylyr rvtsdlrrar rrigeleskd nepiaivgm
 51 crlpggvnsp eslwdlvrsq gdaaisgfpvd rgwdletltg ngdgssathe
 10 101 ggflydaaef daaffgispr eatamdpqqr lllevaweal eragiaptal
 151 rgsrsgvfvq syhwgapsad aatelhghal tgtaasvlsq rlaytlgleg
 201 pavtvdtacs sslvalhlaa qslrvgessl aviggvtilt epsvfvefsa
 251 qgglapdgrc kafsdaadgt gwaegvgvlv aerlsdaqrn ghpvlavlrq
 301 savnqdgasn gltapngpsq erviqqalar tgltpadida veahgtgtrl
 15 351 gdpieaqall atyggqhtpd qplwlgsllks nightqaaag vagvikmvma
 401 lrhghlpptl hadapsshvd wsagsvrlt egqgwpetgr prraavssfg
 451 isgtnahall eqaphpadta dagddaapte pagapaalp ivsgshspqal
 501 rdqaaalaar vetdpalrpq dightlhtar allerravvv apdraellaa
 551 thelaagrsa navvegladv egrtvfvfpg qgsqwvgmga qlldesavfa
 20 601 eriaecaaal aeftdswlvd vlrgvvgaps lervdvvqpa sfavmvslaa
 651 lwgsrgvlpd avvghsqgei aaavvsgals lrdgarvval rsqaigrala
 701 grggmmsval svdvleprlv efegrsvvaa vngprsvvva gepealdalh
 751 arltaddira rriavdyash shqvedlhee llevlaelap rtsevpffst
 801 vtgdwldtar mdagywfrnl rgrvrfadav adllaaeyra fvevsshvpl
 25 851 tmavldliee agvtavatgt lrrdggagr fllsaaevfv rgvdvdwaga
 901 fegtgaarvd lptyafqrer ywntrtaadr tpadapmdae fwaaveqadv
 951 saltaalgtd edsvaailpg ltswrrarsq rttldswryr vtwtplaqvp
 1001 ratltgtwll vttdgiddtd vagalesyga evrrlvldee ctdravlrer
 1051 lagaedvtgi vsvlaaaedd aarhpgltrg laltvslvqa lgdaeatapl
 30 1101 wfltrgafat gpsdpvtrpl qsqiagvgwt talehpqrwg gtvdlpdtld
 1151 araaqrlaaa lsgalgaedq lavraagvla rrivraghra grpartwapr
 1201 gttlitggsg tlapqlarwl aergaehvvl vsrrgadapg apeliaeaae
 1251 sgtevtvaac ditdrdavaa lladltadgr tlrtvihaaa aielsaladt
 1301 tvaefadvvh akvtgarild ellddaeldd fvlvsstagn wsggvhaayv
 35 1351 agnaylsala eqrrarglrt tsihwgkwpd drareladph rirrsogleyl
 1401 dpelaltalq hvldddetvi glmidwdty hdvftagrpa hlfdqipevr
 1451 rrlldqasvpd pagpaadgla arlhglaaa qdrllltlvr teaaavlgla

1501 saesfperra frdlgfdsvt avdlrnrlva gtglrlpstm vfdhpncaal
 1551 aafllkttalg vpgaapqgha atgtpadddp iavigmscry pggatpeel
 1601 lrlaldgadv isefpadrgw darglydpdp drpghtysvq ggflheaagf
 1651 dpgffgispr eavamdpqqr lletsweaf eragidpasl rgsaagtffg
 5 1701 asyqdyssv qngtgeseah mvtgtaasvl sgrvsyllgl egpavtvdt
 1751 cssslvalhl acqslrdges slalaggaav matphafvgf srqralakdg
 1801 rckpfsdtad gmtlaegvgv vllrlshar anghrvlavi rgsavngdga
 1851 snlgtapngp sqqrvirqal anagltgadv daveahgtgt klgdpieaqa
 1901 llatygqdrd aerplllgsv ksnightqaa agvagvikmv lamdagelpg
 10 1951 tlhldapssh vdwtagavel lrgtrpwpep grprragvss fgisgtnahl
 2001 ileqapatep padpdrldt atdtvvpwpl aakspaalra qaarllatve
 2051 hdpdlppapv ghalattraa lehravvge rredflrgla alstgastag
 2101 lvsgiaagdp egavfvpgq gsqwgmgre llatsevftr aiddcatala
 2151 pyvdwslhdv lagegdpall ervdvvpal fammvglal wrshgvvpaa
 15 2201 vvhsggeia aacvagalsl adaarvvalr sqalpqlsgr ggmsvsapv
 2251 ervtallapw gealsvaavn gpssvvsgd tdaldalhta cqeegvrark
 2301 vsvdyashgr hveavrdela rvlavdpdra pevfyfystvt gdrvdadafd
 2351 gaywytnlrq tvrmeeatra llaaghrvfi evsphpvlaa piqetqeava
 2401 eatggsavvl gslrrdegpp rrfltslaea hthgapvdtw ttfarsayqp
 20 2451 vdlptypfqr qdfwpearpa tpaagadasd aafwqlvenq dlaaladalq
 2501 vpaddehtal gtvlpalsaw rakaqartri delryhvqwt rvaepaaapt
 2551 tgrllvavpp dhadapwaa aldalgtdiv rfeakgttra gwaagialv
 2601 edgeeftgfv sllaaedlh pdfgsvplgl gqtlvlvqal gdagltaplw
 2651 cltrgavatg rddaldsptq galwglgrvv alehpdwgg lidlpatlda
 25 2701 raaarltagll adpatedqla vratgvlarr mvhaapsapr tgrrrwrgt
 2751 clitggtggi ggrvarwmae hgaahlvlt rrgpdapgaa alraelealg
 2801 arvtlaacdvd adrdalaall adlpadqplt svfhsagvad gdaraadltl
 2851 dqldallrak ltaahhlhel tapldldafv lfssgaavwg sggqpgyaaa
 2901 nayldalaah rrslldpgas vawgtwgevg matvpevher lhrqgvrame
 30 2951 pdhaigalqq mledddttla vtlmdweafa psftatrpsa lfstvpeavr
 3001 avtgdpgtta gddvdsatpp lrrhleelsa aergralvea vraeasatlg
 3051 htdpdaipag rafrdvghds vtavelrnrl rtalglplpa alvfdhptpt
 3101 alaghlgall fgtapedagt grpddpdari realatvpig rlrkaglldm
 3151 vlkladgdat dapapeadap sesliddmdae allrlatens an

35

SEO ID NO: 7**NysC**

5 1 mstnpdkyve alrsslkeie rlrqneqlv aaavepvavv gigcrfpggv
 51 tspedlwelv aegrdivigpf pqdrgwdlek lagggeggsl aqvvggfveda
 101 agfdpgffgi spreavampd qqrilleitw ealeragidp stlrgtptgv
 151 fvgttgqdyg evikasaedv evysttghaa svisgrlsyt lgaegpavtv
 201 dtgcssslva lhwavqalrg gecsmalagg asimatpgpf vaftaqsgla
 10 251 adgrckpfsd radgtgwgeg agmlvlmrls daqregrpvl avlrqsainq
 301 dgasngltap ngpsqqrvir aaldsahlta adidaveahg tgttlgdpie
 351 aqallatygg drprplwlgs vksnightqa asgaagvikm imalqrgvlp
 401 rslhateptt dvdwtagsvd lldetvawpe tgrarragvs sfgisgtnah
 451 vileqaptap eeptteptvr pavvpwalsa rtaaaldaqr arltghladt
 15 501 pdadpldvgy aladgratfe hravllpdgt elahgtageg pcavlfsggg
 551 sqrpgmgrel harfpvfaaa fdeitalldt hldrplrevv wgtdadllnd
 601 tgwagpalfa vevalyrlva slgvtpdfvg ghsigelaaa hvagvlsled
 651 actlvaarar lmqalprgga mlairatede vtphltddvs iaavngptsv
 701 vvagteeava aigarftaqd rkttrlrsh afhsplmdpm laefravaag
 20 751 ltyhepripv lsnltgtvaa vadlcsadyw vrhvreavrf adgvtaalt dr
 801 gvttlvelgp dgvlsamaqe slpdgaaavp llrkdrpeel savtglarah
 851 vrgvtvrwag lfdgtgarra dlptypfqhq rfwpataaraa qdvtaaglg
 901 adhpllgatv eladgagylf tsrlsvrthp wladhgvqgr allpgtafve
 951 lavragdeag cdrveeltla aplvlpergg vqlqvrvgap daagrtrtigi
 25 1001 fsrvedgfdl pwsqhatgvl tagagapdpt fdatvwppsg aepvdltgay
 1051 erlaalgfgy gpafqglraa wrdtevyae valpdgadtd paafglhpal
 1101 ldaaqhaaay adlgaisrgg lpfawegvsl aaagattvra riapagedtv
 1151 tiavydaagg tvlsvdsllvs revpadapga agtvhrdslf hvewtplqgr
 1201 pgpapatvav lgpdpdalad tlrattgirtt aprdlaalad aegvpvdlv
 30 1251 ttltttppgap vpdaahatta avlalaqqwl addrfadarl vlvtrgatdg
 1301 tdpaaaaagg lirtartenp grfallldlap dtgrdpdpetl atalaashde
 1351 pdlavrgtdv haarlarvpl atepttwpd gtvlitggtg glgavlarhl
 1401 vathgvrhll lasrrgpaad gaddltaelt glgatvhiaa cdvadpaala
 1451 dllgtvpagh pltvvvhtag vvddgvlgs1 tpqrldtvlr pkadaawhlh
 35 1501 eatrhldlda fvlfssvaat lgspgqanya agnafldala arraatglpa
 1551 tslawgpwtq svgmtsslsd ldveriarsg mppltleqgt alfdaalaag
 1601 paalapvrld lpvlrtqgdi apllrglirt pvrtaaqvs qtadglaqrl

1651 agldaaarre allelvrtqi aqvlghadat evetgrqfqd lgfdsltave
 1701 lrnalntatg lrlpatmvfd yptphaladh lrdellgtea esttavpvpt
 1751 rtagtddpiv ivgmacrypg giaspedlwr lvsqgadatg pfptnrgwdl
 1801 dnlydpdpdr pgrthvragg flhdagsfda dffgmsprea matdsqqrll
 5 1851 lelsweaver agidpaslrd sgtgvfagvm yndygttltg deyeafrgng
 1901 sapsvasgrv sytlglegpa vtvdtacsss lvalhwaaga lragecslal
 1951 aggvvmstp stfvefsrqr glapdgrska faeaadgvaw segvgmlvle
 2001 rqsdaavrng h eilavvrgsa vnqdgasngl tapngpsqqr virqalasgg
 2051 lstadvdave ahgtgttlgd pieaqallat ygrdrdpenn lllgsiksni
 10 2101 ghtqaaagva gvikmvmamr hgvlpqtlhv dapsshvds vgavellteq
 2151 tawpetgrar ragvssfgis gtnahvvieq sptavpatpa sadrsvepp
 2201 avpwalsgkt pdalrdqaar llahveahpa lrpvdisysl iatrtafdh
 2251 avvlgtdrae alraltalaa getdpaaltg tvrtgrtafl fsgqgsqrlg
 2301 mgrvlyerfp afaealtdvl taldaelghp lrdiiwgeda qlvdrtygtq
 15 2351 palfaieval frlleawgit pdfvaghsig eiaaahvagv lslgdacrly
 2401 varavlmqsl peggamiavq atedevlpll tddvsiaavn sptsvvsgy
 2451 enatlavarh fadqgrttr lrvshafhsp lmapmlddfr avvesltfta
 2501 pttpvvsnl gelapaealc sadywvrhvr eavrfadgir tladrgvtff
 2551 velgpdsvls amagesapeg agtipllrrd rpeeqavlaa lchlqvlgve
 20 2601 adwsatfrgl dpvrldlpty afqhrfwpa arparpddvr aaglgaaehp
 2651 llgaavqlpd ddgalftgrl slrthpwlad htvlgvtllp gtalvelavr
 2701 agdetgsghl eeltlaaplt lpedgatllq vrvgsaddtg rrtvtvharp
 2751 ddtadrtwtl hatgvlattp paaaafdtv wppadaeplt tddcyahftt
 2801 hrfaygpafq glraawragd vlyaevalpe satdeaaaafg lhpalldagl
 25 2851 haalladdrd tglpfswegv tlhasgatal rvrlapngpn glsvtaadpa
 2901 gnpvatvtrl larpldaeql tihsaltrda lfhlwdtpvp lpdtransapp
 2951 allgpdta vl adalgpava rhatlddlla gdtppatvl vplgapldgd
 3001 taqhahaltr saltlvqqwl atdrldsr l vfvthgavat ddapptdlaa
 3051 aavwglirsa qtenpgtftl lldltepdst talsraltld epqlllragr
 30 3101 araarltrtp apttttthtpw sadgtvlvtg gtgglgglva rhlvrscgvr
 3151 hllltsrsgv gaagaaglva eleslgarvv vaacdvgdgs avelvags
 3201 esyplsavvh aagvlldgvv gsltperlaa vlrvkvdgaw nlheatrgld
 3251 ldafvvfssv agvfggagga nyaagnaflld almvhrvagg lpgvslawga
 3301 wdqvgmtag lterdvrraa esgmplltvd qgvalfdaal atgsaalvpv
 35 3351 rldlaalrtr gdiaplrlgrl vraplrrtaa tglatgadtg lvqrlgrldh
 3401 aqrhealldm vrssaalvlv hadgnaidae rafrdlgfds ltavelrnl
 3451 rtatghlsa tmvfdhptls alaehlrldel fgavesevrv pvqalpptad

3501 dpivvvgmac rfpggvtspe dlwrlvddgt daittfptnr gwldldnlydp
 3551 dpehfgtsyt rsggflheag efdpaffgms prealatdsq qrlllesswe
 3601 aieragidpl tlrsgatgvf agvmysdygs ilggkefegf qgggsagsva
 3651 sgrvsyalgf egpavtvdta cssslvalhw aaqalragec slalaggvtv
 5 3701 mstpstfvef srqrglapdg rskafaeaad gvgwsegvgi lvlerqsav
 3751 rngheilavi rgsavnqdg sngltapngp sqqrvirgal asggilstadv
 3801 daveahgtgt tlgdpieaqa llatygrdrd penplwlgsi ksnightqaa
 3851 agvagvikmv mamrhgvlpq tlhvdapssh vdwsvgavel lteqtawpet
 3901 grvrragvss fgisgtnahv iveqpaves paaepsgrep gvvplplsgk
 10 3951 spealrdqaa rllaglaerp alrpldlgys lattrsafdh ravvlatdra
 4001 davraltala aadadlsavv gdtrtgrhav lfsgqgsqrl gmgrelyerf
 4051 pvfaealdva idhlldaalpa gaslrevmwg ddvelldetg wtqpalfave
 4101 valfrlvesw gvrpdfvagh sigeiaaahv vgvfsledac rlvaaratlm
 4151 galptggami aigaaedevt qhltddivsia avngptsvvv sgaesaartv
 15 4201 adrlaengrk ttrlrshaf hsplmdpmla efravaegls yatptlpvvs
 4251 nltgrlatad dlcsaeywar hvreavrfad gvstlenegv ttfllelpgdg
 4301 vlsamaqqsl tgdaatvpal rkdrdeetsa ltalahlhta glrvdwaaff
 4351 agsgatrvidl ptyafqhaty wptgtlptah aaavgltaae hpllngsvel
 4401 aegegvlfth rlsiqshpwl adhavgqvl lpgtallela fragdeagcd
 20 4451 rveeltlaap lvlpergavq tqvrvgvadd tgrrtvtvhs rpehatdvsw
 4501 tqhatgtltm gsapadtgfd atawppadae platddcyar fttlgfaygp
 4551 vfqqlraawr agdvlyaeva laestgdeat afglhpalld aalhaslvah
 4601 egeesnggpl fswegatlya tgatalrvrl tptgtdgrsv aiavadtagr
 4651 pvaaaidnlvs rrvsgdqltg aaglardalf tldwnvpven lvpenvpven
 25 4701 tggghaqdqd grpaaatval vgadgtaiaa dltaagihtt lhpdlttlat
 4751 tdadvpktivl ipltgtgtgt gtgtestdgi gtgaaesdas apspaevaht
 4801 lstaalalvq ewtaqerfag srlafvttga taaggtdvmd vaaaavwglv
 4851 rsaqseapdt fvlidrdpgp agthdrtaaa ergqlllral htdepqlalr
 4901 dggvlaarla rfdtaaaltp padrawrlds takgsln gla ltpypaalap
 30 4951 ltghevrvev raaglnfrdv lnalgmypgd dvgsfgseaa gvvvevgpev
 5001 tglapgdqvm gmitgsfgsl avddarrrlar lpedwswetg asvplvflta
 5051 yyalkelggl ragekvlvha gagvgvmaai qiarhvgaev fatasegkwd
 5101 vlrslgvadd hiassrtldf eaafaevagd rgldvvlsl agdfvdasmr
 5151 llgdggrfle mgktdiraad svpdglisyqs fdlawvvpel igtmlaelmd
 35 5201 lftrgalrpl pvrtwdvrha kdafrfmsma khigkivltl prswkpegtv
 5251 lvtgggtgglg glvarhlvrs cgvrhlills rsgvgaagaa glvaeleslg
 5301 arvvaacdvd gdgsavaelv agvsesypls avvhaagvld dgvgvsltpe

5351 rlaavlrpkv dgawnlheat rgldldafvv fssvagvfgg agqanyaagn
 5401 aflaldmvhr vaggldpgvsl awgawdqgvg mtaglterdv rraaesgmpl
 5451 ltvdqgvalf daalatgsaa lvpvrldlaa lrtrgdiapl lrglvkapir
 5501 raaattpgdt glaeqltrlq raerrdtlla lvrdaqamvl ghtsgdgvdp
 5 5551 srafrdlgfd sltavelrnr igaatglrlp atavfdypta dalaahlte
 5601 llgpdaesdp depgdptagp tddpiviigm scrfpgdigs pedlwrllgd
 5651 gadvvtdfpt nrgwdldnly dpdpahagts yartggflhd aadfdadffg
 5701 mspreamatd sqrrllless weaieragid pltlrdsrtg vfagvmysgy
 5751 gtrldgaefe gfqgggsals vasgrvsytf gfegpamtvd tacssslval
 10 5801 hlaaqalrgg ectlalaggv tvmsipdtfi efsrqrglap dgrskpfses
 5851 adgvgwsegu gmllderqsd avrnghqila vvrsgavnqd gasngltapn
 5901 gpsqqrviro alasggldsta dvdaveahgt gttlgdpiea gallatygrd
 5951 rdpenplllg sikslnlghtq aaagvagvik mvmmamrhgvl prslniteps
 6001 shvdwsagav ellteqtawp etgrarragi ssfgisgtna hvileqpeaa
 15 6051 rhsapeeadt aeaaakapat ahlpvmpwal sgktpealra qaarllahlq
 6101 qrpelapadi alslatqrsq fthravvlst drdeatrls alattaasdp
 6151 saltgtvtmg rcavlfsggg sqrlgmgrl yerfpvfaea ldvvidhlida
 6201 alpaqaglr vmwgddvell netgwtqpal faievalfrl veswgvrpdf
 6251 vaghsigeia aahvvgvfl edacrlvaar atlmqalpag gamiavgate
 20 6301 deviphltde vaiaavngpt svvisgaea tqtvaqhfad qgrrttalrv
 6351 shafhspimm laefravaeg lsyatptlpv vsnltggvat adelcsaeyw
 6401 vrhvreaurf adgvtaleae gvrftflelqp dgvlaamare tvaddtvtvp
 6451 vlrrnmpeer tlltalgrlh ttgtpidwaa llaptgarpv dlptyafqhr
 6501 pfwpsgprdt adaaavgiag ashpllngiv eladeegllf tgrlsllqshp
 25 6551 wladhavmgq vllpgtalle lalragdevg cdhveeltla aplvlperga
 6601 vqtqvrvgva dttgrrtvti hsrparattt dsdthtgdtd pwtqhatgvl
 6651 vaglpatatv pfdatvwppa haepvdladf yasragegfg ygpafqglra
 6701 awrrdgevfa dvalpeagrt eaeayglhpa lldagllaaw lvapdgeptr
 6751 tgsvpfswrg vflaasgass vrurlgrdsd gtllslaiadt tgapvasvqa
 30 6801 lsmrtvsvta lsataglard alfrldwasa pepacqpddt vtvipavavv
 6851 gtetseltse ltaalraaga dvdvrttlst depapalial plvasdqtgt
 6901 aeaapvpaav hdltrralal vqtrlqeqhf adtkfvfvtr gatvgrdvaa
 6951 aavwglvrsa qsenpgcfal vldldpdgav aaalvaalvs gepqlavrgd
 7001 vlrvarlvrr pltevgagad gtgdgvgggs gvsfsgegav lvtgggtgglg
 35 7051 avlarhlvae ygvrldllls rsgervagag elvaelagvg arvrvvacdv
 7101 tdraavvelv gghavsavvh aagvlddgmvt galtgerlsa vlrvpkvdavw
 7151 hlheatrgld ldafvvfssl agvfgspgga nyaaanafld almtrrraeg

7201 lpglslawgp wsltdgtsgm ladaeadrlt rsgvppltae qglalfdaal
 7251 atgdatcvpv rldlsalraq gevppllrsl irgrsrraaa aesatatglr
 7301 erlvglnpve rqevlldlvr gqvalvlgha daddvhpara frelgfidslt
 7351 svelnrnlnt vtglrlpatm vfdyptvevl vsyvldellg tdaevatvqp
 5 7401 aavavaddpi vivgmacryp ggvaspddlw rlvtgdvdav spfptnrgwd
 7451 veslyhpdpd hlgtsytrsg gflheagefd pgffgmspre alatdsqqr1
 7501 llessweaie ragidpvs1r gsrtgvfagv mysdysamla spfefegfggs
 7551 gsspslasgr vaytlglegp avtvdacass slvamhwamq alrsgecgla
 7601 laggvvmst pavfvdfarq rglspdgrck afadaadgvq wsegvgvlvl
 10 7651 erqsdavrng heilavvrgs avnqdgasng ltapngpsqq rvirqalasg
 7701 gltagdvdiv eahgtgttlg dpieaqalla tygrdreper plllgsvksn
 7751 lghtqaaagv agvikmvlam rhgvvprtlh vdapsshvdw segavellse
 7801 qaawpetgrv rragvssfgi sgtnahvile rpeaarrpvm etntvepstv
 7851 pwvlsktpc alraqaakll ssieerpelr lvdvgmslt grstfehrav
 15 7901 vlaadradaa ralsaiiaade adaaaatgrv gagrhavlf s gqgaqrlmg
 7951 relyerfpvf aealdvvvdh ldaalpaqag lrevmwgdda ellnetgwtq
 8001 palfaieval frlveswgv pdfvaghsig eiaaahvagv fsledacrlv
 8051 aaratlmqal paggamiaqv atedevtphl tddvaiaain gpnalvsgv
 8101 edaaveigar faaegrtrtr lhwshafhsp lmdpmlaefr vvaeglsyaa
 20 8151 pslpvvsnl1 gqvataclc saeywvrhvr eavrfadgt aleaegvrtf
 8201 lelpgdgvla amagasltes slavpllrkd rpeepaalaa laqlhiagar
 8251 vdwplvfagv gagrvelpty afqrgwfwpv grvgvggdvg avglgsaghp
 8301 llgaavelaa gagvltgrl slsshgwlad havmgrvfv gtallemvmr
 8351 agdevgcgrv eeltlaapl1 lperggvrvg vavdapdaag rrgvgvyscp
 25 8401 dgvgqavwsq havgvlasgv adqvggfgdg gvwppqgavs vdaegcyelf
 8451 adagfgygpv fgg1ravwrr geelfaeval sdevaesadt atgfglhpal
 8501 ldaslhasll sslegqsadg gpalpfaweg vs1fasgata lrvrlapage
 8551 havsvtavdp tgapvisida lrtrrltlde vnashtqlsd alfgvqwttv
 8601 pstpaadhps vaiigtdhlg laealssssa gattttttaa yesldaliaa
 30 8651 gpevsvpdvt liglttedai aqyvndhdat vagqgtigag aaavdaarrl
 8701 taealrtiga wladerlaar rlvfvtrgaa dgqdvaaaav qglvrsaqte
 8751 npgtfglldl dgteastavl gealtsdepq lllrdghlha arltrlaspa
 8801 dtavptewna dgtvlitgt gglgaqfarh lvdrygvrn1 llvsrrgpda
 8851 pgttelvacl tahgaevavq acdvadgdav aalvagvpde hplravvhta
 35 8901 gvlddgvig1 lteerlatvl rpkadaawhl heatrgldld afvvfssvag
 8951 vfggagqany aaanafldal maqrraaglp glslawgpwd qtggmtgmls
 9001 daeadrlars gipplsaeqg lalfdaalal agtstpdraa gsaaastsqt

9051 gdtiaipaaa lvapvrldla alaaqgevpa ilrglvtrtr rrrtaaggsvt
 9101 vaglvnrslg ltaderrqel lelvtqaal vlghadpasv dstaqfrdlg
 9151 fdsltavelr nrlstatglr ltatlvfdyp ntdalaehlr delfgavese
 9201 vrvpvqalpp taddpivvvg macrfpggvt spedlwrldv agtdaittfp
 5 9251 tnrgwdlesl ydpdpahlgf sytrsggflh eagefdpaff gmsprealat
 9301 dsqqrllles sweaieragi dpltlrgsat gvfvagmysd ygsilggkef
 9351 egfqqggsag svasgrvsyt lgfegpavtv dtacssslva lhlaaqalra
 9401 gectlalagg vtvmstpgtf vefsrqrgla pdgrskafae aadgvgwseg
 9451 vgilvlerqs davrngheil avirgsavnq dgasngltap ngpsqqrvir
 10 9501 qalasggfst advdaveahg tgttlgdpie aqallatygr drdpenplll
 9551 gsiksnlght qaaagvagvi kmvmamrhgv lpqtlhvdap sshvdwsvga
 9601 vellteqtvw petgrvrrag vssfgisgtn ahvileqpea vqrlapgaee
 9651 tvepvaijps aepslvpwal sgkspealra qaarlrdfia erpeprside
 9701 ghslavtrsq fdhraivldv dakapadsia alaaasgva dpavvsdavs
 15 9751 tggsaviftg qgaqrlgmgr elygrfpvfa ealdvvvdhl daalpaqagl
 9801 revmwgddve llnetgwtqp alfavevalf rlvewgvrp dfvaghsize
 9851 iaahvagvf sledacriva aratlmqalp tggamivqa tedevtpht
 9901 devaiaavng ptsvvisgae eatqtvaqhf adqgrrtal rvshafhspl
 9951 mdpmlaefra vaeglsyatp slpvvsnlgt wlatadelcs aeywvrhvre
 20 10001 avrfadgitt leaegvrtfl elgpdgilsa laqqslagea vtvpvlrkdr
 10051 geestaltar ahlhtrglie dwqdfagvg agrvelptya fqrgrfwpvvg
 10101 rvvgvgdvga vglgsaghl lgaavelaag agvvtgrls lsshgwladh
 10151 avmgrvfapg tallemvmra gdevgcgrve eltlaaplvl perggvrqv
 10201 avdapdaagr rgvgvyscpd gvgqavwsqh avgvasgaa dqvggfgdgg
 25 10251 vwppqgavsv daegcyelfa dagfgygpvf qglravwrrg eelfaevals
 10301 devaesadta tgfglhpal daslhaslls slegqsadgg palpfawegv
 10351 slfasgatal rvrlapageh avsvtavdpt gapvisidal rtrrltldv
 10401 nashtqlsda lfgvqwtvp stpaadhpsv aiigtdpfgl adglldalpl
 10451 veergdlaal aasehpvpdl vlvpvagtrr tgvpadaegh tdagtsdmlr
 30 10501 svreataqvl eqiqqladd rfeaarlvfv trgavsvgeg giadlaasav
 10551 wglvrseqse npgcfglldl dldlaldsl apevdierdr drdpvggtvg
 10601 palaaalhat adepqlalrg gtvqaarltr ipapqtdrae tdaetdrpe
 10651 idtrrpgtvl itggtgglgg llarhlvaer gvrslvlasr sglaaegaek
 10701 lvadlealga vvavqtcvda dgdaaalva gvsdeyplta vvhtagvldd
 35 10751 gvigslteer latvlrpkad aawhlheatr dldldafvvf sslagvlgga
 10801 gganyaaant fldalmaqrr aaglpvgsia wgpwdraggm tglldaead
 10851 rlarsgvppi saeqglalyd aatagerplv vpvrlldaal rglgdvpall

9051 gdtiaipaaa lvapvrldla alaaqgevpa ilrglvtrtr rrrtaaggsvt
 9101 vaglvnrlsg ltaderrqel lelvtqaal vlghadpasv dstaqfrdlg
 9151 fdsltavelr nrlstatglr ltatlvfdyp ntdalaehlr delfgavese
 9201 vrvpvqalpp taddpivvvg macrfpggvt spedlwrlvd agtdaittff
 5 9251 tnrgwdlesl ydpdpahlg t sytrsggflh eagefdpaff gmsprealat
 9301 dsqqrllles sweaieragi dpltlrgsat gvfvagmysd ygsilggkef
 9351 egfqqggsag svasgrvsyt lgfegpavtv dtacssslva lhlaaqalra
 9401 gectlalagg vtvmstpgtf vefsrqrgla pdgrskafae aadgvwseg
 9451 vgilvlerqs davrngheil avirgsavnq dgasngltap ngpsqqrvir
 10 9501 qalasggst advdaveahg tgttlgdpie aqallatygr drdpenplll
 9551 gsiksnlght qaaagvagvi kmvmamrhgv lpqtlhvdap sshvdwsvga
 9601 vellteqtvw petgrvrrag vssfgisgtn ahvileqpea vqrlapgaee
 9651 tvepvaikps aepslvpwal sgkspealra qaarlrdfra erpeprside
 9701 ghslavtrs q fdhraivlvd dakapadsia alaalasgva dpavvsdavs
 15 9751 tggasvlftg qgaqrlgmgr elygrfpvfa ealdvvvdhl daalpaqagl
 9801 revmwgddve llnetgwtqp alfavevalf rlverwgvrp dfvaghsize
 9851 iaahvagvf sledacrlva aratlmqalp tggamiavqa tedevtphlt
 9901 devaiaavng ptsvvisgae eatqtvaqhf adqgrrttal rvshafhspl
 9951 mdpmlaefra vaeglsyatp slpvvsnlgt wlatadelcs aeywvrhvre
 20 10001 avrfadgitt leaegvrtfl elgpdgilsa laqqslagea vtvplvrkdr
 10051 geestaltar ahlhtrglie dwqdfagvg agrvelptya fqrwfwpgv
 10101 rvgvggdvga vglgsaghl lgaavelaag agvvltgrls lsshgwladh
 10151 avmgrvfapg tallemvmra gdevgcgrve eltlaaplvl perggvrqv
 10201 avdapdaagr rgvgvyscpd gvgqavwsqh avgvlasgaa dqvggfgdgg
 25 10251 vwppqgavsv daegcyelfa dagfgygpvf qglravwrrg eelfaevals
 10301 devaesadta tgfglhpall daslhaslls slegqsadgg palpfawegv
 10351 slfasgatal rvrlapageh avsvtavdpt gapvisidal rtrrltldev
 10401 nashtqlsda lfgvqwtvp stpaadhpsv aiigtdpfgl adglldalpl
 10451 veergdlaal aasehpvpdl vlvpvagtrr tgvpadaegh tdagtsdmlr
 30 10501 svreataqvl eqiqqwladd rfeaarlvfv trgavsvgeg giadlaasav
 10551 wglvrsaqse npgcfglldl dldaldsdl apevdierdr drdpvggtvq
 10601 palaaalhat adepqlalrg gtvqaarltr ipapqtdrae tdpaetdrpe
 10651 idtrrpgtvl itgggtgglg llarhlvaer gvrsylvlasr sglaaegaek
 10701 lvadlealga vvavqtcdfa dgdavaalva gvsdeyplta vvhtagvlld
 35 10751 gvigslteer latvlrpkad aawhlheatr dldldafvvf sslagvlgga
 10801 gqanyaaant fldalmaqrr aaglpvgsia wgpwdraggm tgtlsdaead
 10851 rlarsgvppi saeqglalyd aatagerplv vpvrlldlaal rglgdvpall

10901 rglvrtparr taaagaapsa dvltrqlagl ggaeeqevll rlvrgqaavv
 10951 lghadgsaig agrqfqelgf dsltavefrn rlmaatglrl patllfdypt
 11001 padvvghlrg rltggevsga gsvlaaldnl eaviaglsld dagehqlvag
 11051 rlevlrakwa dmrsaegavd ggadvdieea sdddmfalld delgln

5

SEQ ID NO. 8

NysE

10 1 mttsteeslw arcfhpapaa pvrllcfpha ggsasfyfpv
 41 saqlssvaev faiqypgrqd rrkeagvsdl atladqvyda
 81 lrpllkerps tffghsmgat lafevarrfe addgdvlrflf
 121 asgrrapsrv reeavhrrsd dgiveelkll agtntallgd
 161 eeilrmilpa irsdyqaiet yrcppdvtvr apltvlgtgdr
 15 201 dpktsldeae awrghttgdf dlkvlpgghf fvsseapaii
 241 dllrahlagn g

SEQ ID NO. 9

20 **NysR1**

1 mrkqsgssgl lttlvgrdde lrtlarhaaa ardgraglvl lhgpagmgt
 51 sllrsftasd vcrgmtvlyg tgetvagag yggvrellgg lglsggdarr
 101 splleglaar alpaltadpa gpdaatgayp vlhglywlaa rlmaqrplvl
 25 151 vlddvhwcde rslawidfl1 rraedlpll vlawrseaep vapavladia
 201 aqrrptvlg1 hplgpddige mvrrvfrtta apsfvsrvaa vsggnplala
 251 rllldelraeg vrpdaagerr aaevgshvla rsvrcllerr ppwvrgvara
 301 iavlgtpecte llaalagvpa atvdeallvl rragilaadr vdfvhdvrs
 351 avlddvappt laelrtnaal llsdagrpse elagqlml1p vldqpwwaav
 30 401 lrdaaaqaes rgapeagvrc lyrvlevepd nvavriqmar alaeinppea
 451 mrl1keals1 agdvrtraqv avqygftcla vqespsgvrml edalaelta
 501 elgpepgpvd relrtlvesv llivgadekv tigavrdraa rltmppgdt
 551 aqrqmlamtt vltamdgrda rsavdgarra lrapgvelep wlllsasfal
 601 sladevadaq yaldmlqyg qdnaavwtyv lalstrallh hgvga1peal
 35 651 adaqtaveil geerwadgav lprvalatal vdrgeperae hvldgitrpr
 701 lerfvieyhw ylqarayarw vrgdfqgald lllacgrsle esrfsnpafv
 751 pwwadgavll atldrh1dgar elaaygsela erwgtarglg lafmaaggvaa

801 pgragidhlt eavslldasp aramearael llghahlkrd dlraarehrlr
 851 aaadlaqrcg avklgvdark llvtaggrvr rmtaspldml tgmertvadl
 901 avtgasnrai aealfvtvrt iethltsvyr klvgvgrael savletrtat
 951 sgrqppawvs qargra

5

SEO ID NO. 10**NysR2**

10 1 vprskarnqp ttctpqcapd ahgdptmle cgregrligd llhrlgqgrp
 51 svlsltgrpq haqnalvrwg acrarhdglr vlraqatpae relrygavlg
 101 llavldgphg stldaaairhd gppplpvpgi eevlrrtgta ptlvvvedvq
 151 wldpasltwl qillrhlgpd tplavlassc gdttafdtdp kapavpgppd
 201 tvpvarfvvp altdrgvaat vravcgtpgd eefiaaltsa tagnpailrd
 15 251 alrafvdhgl padadhlpel haltagvvgd htvraldglp aevnavlral
 301 avcgdlldfh rvralagahs lsedrirtll asvgltvsvg dkhvirfpas
 351 karviedmpa aeradlyvra aelthscgvn dedvahlllr sspigapwv
 401 pllrrgfaaa lrredhhrac aclsralsqep ldprersllt lelaaaaeva
 451 rpeagdrirlg elvrstvadt dptssgegvg vraidlgfar gnsewvrrta
 20 501 gealpyagpa dreelvalfw laavrdddap mipvvprlpd rpvpapaqaga
 551 rawqlatage dadkarklar ialtggvnes lmpkplaaca alfatddnde
 601 avhgltdmlt aarsahlrsm aarifnlrar ihlcaarlea aerdldsae
 651 alpptswhpr alpnliatri lvsmetgrp d rarrlaeapv paggeegvw
 701 palllararv aaddgdweea lrlsrecgrw lfrhwanpa mlswrplaae
 25 751 aclklgdvte arrlrdeelf fadrwgtasa rgialrttrr lfdddgdvav
 801 rrireaaall rdsparlayl wsrlsqagae tahgdtaaaa rswqavarmt
 851 aahpasrlat aartltvpsv pvatapptav vppgwrldse aekdtvllaa
 901 rghgnrqiae qlavsrtrve lrlsnayrkl riggrkelyl llealegpva
 951 das

30

SEO ID NO. 11**NysR3**

35 1 mllerenela riraaldaae agdsslllin gplgsgrsal lrripelagd
 51 gtrvlrasaa wrerdfpfgi arqlfdhlls gaggagpaer tagaehfsrl
 101 mdtgdrptgt gpalevsqav lqgaqallad asaerrllil vddlqwadgp

151 slrwlahltr rlhglrallv ctldgdhrg ryplvrevag aahtvlrlap
 201 lsrdatrlll agpqgrppqd alvravyeas rgnplfltaf rsalratgrp
 251 pggdhfgavr elsptvlrdr laghlriqqp pvrevavava algdhsdpvl
 301 laqlagvdei gfagarralv dagllargrd vrfvhgvvrd avdsltldde
 5 351 rershddaad llyrcgrpae qvaghllavv hpgrpwseav lrsaahnalr
 401 agrpadaary lrrallhhrt qdgcrarilv dlataerald pdacvrhvsq
 451 avalldtsrd raaavlrripp sllaapspsa velvrqaaag ldepqgrdee
 501 gadelalrle awlrhsghe pvelassvar lrrmgarppv dsvaerelva
 551 vllsagalsg rlsaaeiadt gnrilerepa taahahtplp lvmlslfvae
 10 601 svqgvaswla seqhtrrrya tgaddvllta eraflvltqg rpaaarehve
 651 ralvmdagdw sepavmmfaa vafelrdpal serilerird rrpaglalra
 701 tggmlqaavd vhfgrgrdal dtllacgrrl etvgwrnsal lpwrpyaigl
 751 hqrlgetdaa lqlaedelrw arewgattnl gralrlkgwl lqdegldllr
 801 esveilrass yatelartlv vlgrrlpggp eaeavlreaa giaaacgvvpw
 15 851 laeraelglg saivppvatl tpserrrvasl vsrgltnqai atelgvssra
 901 vekhltsayr klgvsgrrel vnalpgr

SEO ID NO. 1220 **NysR4**

1 visaqtapag esvgpglmas ldrdltikha ngefrrrfdd sagdvcgrsf
 51 rdlmhpsvqq plmrqfsrli egkrhrfash vvavgaqdaa fagtlrtasav
 101 tgktptdiagi lvlmdssgaa daadagvvtg qkkflteida rilegiaagl
 25 151 stiplasrly lsrqgveyhv tgllrklrvp nraalvsray smgilnvgtw
 201 ppkvvddfik

SEO ID NO. 1330 **NysR5**

1 vdaegrerrdm lelirrsgsa dvvrlaeefa vsketvrrdl nvlegghlir
 51 rrrhggaypmv rpgseavfvs rtaqpipees riataaaell seaetvfide
 101 gftpqliada lprdrpltiv taslpvvsaf atspqanvll lggrvrrgtt
 35 151 atvdhwavhm lsgfvidlaf lgaegisrry glttdpava evkaqairva
 201 rrpvlagvht kfgtasfcf gevgdletiv tgaglpvaea hryhlmgpkv
 251 lrv

SEQ ID NO. 14**ORF2**

5 1 vaqdsqgtpr sldhvdqalv halqitpras wtrigsvlgl dautvarrwn
 51 rlvetgaawi schpavlaa sgqgclafve idcapgrlld varalaaaph
 101 vvalshvtgd rdlqlnvmar dpamlsrwtv hdlaaldgvr aarthlagpv
 151 htegsrwrlr algrhqvarl aadasrhrtv tpaflvdeld qqlvtalsvd
 201 gratyrallae qcgagpdtvr rrvqrlfaad mlharcevar plsewptvts
 10 251 fwgqvpaarl revtrrvtgm revrlcasvi srhnlhlvaw vrslddaqrf
 301 evrlaeraad ltvteraval whmkhgghll deegyrvgtv plalwreptd
 351 arrg

SEQ ID NO. 15

15

ORF1

 1 mqpelerlrr slhrepelgl alprtqekvl aaldglplei tpgtgltsvt
 51 avlrgrpgg avllrgdmda lpvteesgvp yaseipgrmh acghdlhtag
 20 101 lvgaarllae rrerlhgdvv fmfqpgeegh ngagamieeg vldaagkpdl
 151 aayalhvasn qvpggvitr sgtitsasdl ltvtrvgegg hgstpatakdl
 201 pvpaacemvt avqnwvtraf difdpvvvtv gtfhagtkas vipdnrvpg
 251 hgaqlfrirp gparrrelpr lvraiaaahg leaevdyllrh ypvtnvdtde
 301 tafavatard lfapgevves piptngsedf ayvlrrvpga mllvgaapeg
 25 351 sdwqhaphmnh spravfddrv lyrqaallte laarraata paepavag

Translation products of SEQ ID No. 2 (Nys 2):SEQ ID NO. 16

30

NysF

 1 vielilpatv ateaayddrp rpgdrllsse reviaraves rgreftttrh
 51 larralrrlg hpdrailpnr rgapqwpvgi vgsmtbcagy raaavspael
 35 101 saavsideap ngplpagvln aialpserph lvalaahrpv vhwdrillfsa
 151 kesvfkawyp ltqreldfse aeividptgg aftarllvpg pllggrrvtv
 201 fpgrwhstpa llttavhlpa ptprrdrehr thltvnsplp rptfg

SEQ ID NO. 17**NysG**

5 1 maspddleee rtaprrparrl vglrrphrrs valavsmgvg givlnafgpl
 51 llgrvtdlia dglvggvpqp apgidfaaig rlllvllaly vvaslfmlaq
 101 grlvasavwr tihelrrdar ekltrlplrh fdrqpagell srttndidnl
 151 qqtllqqtlae litsifsllt mlvlmlvisp slavvmlsv pvsaliaari
 201 skraqphyaa qwsangtlina hveevctgha likgfdrraa aeerfdacnd
 10 251 avyraaakaq fasgamepvm mfvanlgyvl vavigawkvi ngltltlgdvq
 301 afilyarqfs qpiveiasva grlqsgiasa qrvftlldap eqapdplrpg
 351 tparaegrve ftdvsfrysp dtplienlsl tvepgstvai vgptgagktt
 401 lgnllmrfe pdsgrilldg tdtatmtrdd lrsrfglvlg dtwlfggtia
 451 eniaygapga cradieear atcadrfirt lpqgydtvld desgtvsage
 15 501 kqltvaraf larpavlvld eatssvdrtr evliqramns lragrtsfvi
 551 ahrlstirda dlivmdagr iveggthdql lcaqglyarl haarthtpta
 601 gaaag

SEQ ID NO. 18

20

NysH

 1 vllrllraql rpyawataal valqlvqilg tlllptlgaa lidqgvvrgd
 51 ggritelgvv mgvvalvqia aalgaaalaa rtatamgrdl rsalfrrild
 25 101 fsareigrfg tpslltrsvn dvqqvqnlq tgfgivvcap lmcigsvlla
 151 lrqdvplall lvalvlvav cfglllarmg tlyarmqltl drlgrllrea
 201 itgvrvvrsf vrddherarf aqtndaflvv srrvgrliat mlpvvlllmn
 251 gftvallwtg shridagrmv igsalsallsy lslilmsvmm lafvflsvpr
 301 arvcagriae vldtgssvap paapqpvrpg agrielcaag yrypgaeepv
 30 351 lrdvdltevpe geriavlgst gsgkttllnl vlrladateg avrvvgtdvr
 401 eltaatlaaa vgfvpqrpyl fsgtvasnlr fgrpdatdee lwealrvaqa
 451 adfvarmpdg ldaeitqggg nvsggqrqrl slarallrrp eiylfddcfs
 501 aldqatdaal rtalvpytag atvitvaqri sagrdadriv vldrgrvvaq
 551 gthdvllrts ptyreialsq lteeeeahgl agrp

35

SEO ID NO. 19**NysD3**

5 1 mskralitgi tgqdgssylae hllsqgyqvw glirgqanpr ksrvsrlase
 51 ldfidgdlmd qgslvsavdt vqpdevynlg aisfvpmwq qaelvtevng
 101 mgvlrmleai rmvsglstsr tvsprgqirf yqassemfq kaaetpqret
 151 tlfhprspyg aakayghyt rnyresfgmy avsgmlfnhe sprrgqefvt
 201 rkislavari kqglqdklal gnldavrdwg yagdyvramh lmlqqdagdd
 10 251 yvigtgqmhs vrdavriafe hvglnwedyv vidpdlvrpa evevlcadsa
 301 kaqdrlgwkp dvdfptlmrm mvdsdlaqvs renqygdvll aanw

SEO ID NO. 2015 **NysI aa sequence (partial)**

 1 mdneqklrdy lklatadlrr trrrvhkles aaqepvaiig mtcrypggvr
 51 spedlwrme agehgvtpfp tdrqwdleal aaaptasggf lhdapdfdad
 101 ffgispreav amdpqqravl esaweaferra gidptsvkgs rtgvfigama
 20 151 qdyrvgpapg aegfqltgnt gsvlsgrisy tfgtvgpavt vdtacssslv
 201 avhlatqalr agectlalag gvtimsgpgt fiemgrqggf sadgrcrsfg
 251 dtadgtgwae gvgilvlerl sdavrnghei lavvrgtavn qdgasnglta
 301 pngpsqqqvi qqalvnarla agdidvveah gtgttlgdpv eaqallatyg
 351 qnrpadrp11 lgsvkns1sh tqaaagvagv ikmvmamrhg t1prtlhaee
 25 401 pthhvdwsqg avrlltdttd wpatgaprra avssfgisgt nahtiieqap
 451 epqpedaata qddaagstpa tapvvpgvvp vllsgrtpda lrgqaaalra
 501 aldtgrprdl ldla1slatt ragfehravl latdhpaltd gltaladadd
 551 paaapawitg ttraetrlav lftgqgaqrl gagrelaarf pafataldaa
 601 ldaftphldr plrevlwgt1 aalldrtaya qpalfaveva lyr1iesfgv
 30 651 rpdhlaghs1v geivaahlag v1sladaatl vaargrlmq1a lpdggamiav
 701 qaseadvapl laghedqvai aavngpsavv lsgaeatvta laeq1aadgr
 751 ktrrlrvsha fhsplmepml dafravvedl t1qppllpv1v snltgkpatv
 801 aqltsadywv dhvrhavrfa dgidwlarhd ttaflel1gpd gvlsamaqdc
 851 ldaadadavt lpalragrpe ehtl1ttalag l1hvhgatldw tgc1fagtgar
 35 901 rtdlptyafq rrrywpkalq sgtadlrs1v lgaahhp11s aavsladagg
 951 t1ltgr1srq thpwladhtv rgtt1lp1gta fl1elavragd evgcdrveel

1001 tlaaplllpe qggvqvqlwi gnpdvsgrrt vnvharptdg ddtptwahat
 1051 gvlttadasr qlpasseqgg tplagdphpa ldaaqwppag aeplpldghy
 1101 drladggfgy gpvfqglraa wrggdvvyae velpeagrsd aeafglhpal
 1151 ldaalhaapf tglgergrgg lpfswegvsl haggattlr rltpvaddal
 5 1201 altvadgtga pvlsvdsllv rsvatqqltd aaavardalf rldwtpvqpt
 1251 atdpgpvall gadpfgllth agfadapayp dlaalaaadg pvpttvvls
 1301 agtgddaadp arsahrcae alaavqtwld hherfaaarl vfvtrgatvg
 1351 rdvaavawg lvrsaqsenn gcfalvdldp dgavgaalv aalvsgepql
 1401 avrgdvlrva rlvrrpltev gagadgtgdg vgdsgsvsfs gegavlvtgg
 10 1451 tgglgavlar hlvaeygvrd lllvsrsger avgagelvae lagvgarvr
 1501 vacdvttraa vvelvgghav savvhaagvl ddgmvgaltg erlsavlrpk
 1551 vdavwhlhea trglldlafv vfsslagvfg spgganyaaa nafldalmtr
 1601 rraeglppls lawgpweqsg gmtgtltdvd aerlarsgvp plsvaqqal
 1651 fdaavagtda tcvpvrldlp vlrangevpp llrslirvra rraavagsat
 15 1701 agnlaqlrr ldedgrdemv ldlvrgqval vlghatggdv dagrafrldg
 1751 fdsltavelr nrlntvtglr lpatlvfdyp tvrhlatyvl dellgtdae
 1801 atvqpaavav addpivivgm acrypggvss pedlwrvtte gtdavsgfpt
 1851 nrgwdvesly hpdpdhpgts ytrsggflhe agefdpgffg msprealatd
 1901 sqqrllless weaieragid pvslrgsrtg vfagvmysdy samlaspefe
 20 1951 gfgsgssps lasgrvaytl glegpavtd tacssslvam hwamqalrsg
 2001 ecglalaggv tvmtspavfv dfarqrglsp dgrckafada adgvwgsegv
 2051 gvlvlerqsd avrnghela vvrsgavnqd gasngltapn gpsqqrvirq
 2101 alasggltag dvdvveahgt gttlgdpiea qallatygrd reperplllg
 2151 svksnlgthq aaagvagvik mvlamrhgv prtllhvdaps shvdwsegav
 25 2201 ellsegaawp etgrvrragv ssfgisgtnv hviveqapga kaiaaagaar
 2251 rtpgavpvl sgrgrsalrg qaarllghlq arpdaelvdl alslatrstr
 2301 feqraavvaq drdqliaslg alaadrpdpa vvegeaagrg rtavltftggg
 2351 sqraamgrel hevqpefaaa fdavcavfdp lldrplrevv faedgsdeaa
 2401 lldetgwtqp alfavevalf rlveswgvrp dfvagsige iaaahvagvl
 30 2451 tledacrlva aratlmqalp tggamiaiaq tedeiaahld dtvaiaavng
 2501 pqsvvisgde eaaetiaatf aergrktrkl rvshafhspr mdgmldafri
 2551 vaegltyrp riplvsdltd rraddaevct aeywvrhvre avrfadcvrt
 2601 lrdagattfl elgsdgllda maedtlgddh daelvplmlra graeelaat
 2651 alarlqvrqvdv dwaaaylag tgarrrtdlpt yafghayywp qlptpaaala
 35 2701 aadpadqqlw aavergdare ladilglgeq dltpldsllp altswrrngq
 2751 ekhlldtlry rvewtrlskp tapvldgtwl lvasdataad qpalldglad
 2801 algshgarvr rllldscad ravlaerlar tadvdaatqv lsvlplderd

2851 addcppltrg laltvalvqa ladtgaqgrl wtatrgavst npadpvthpv
 2901 qaaawglgrg valehprlwg glvdlpqvfd eragqrlagi lavkdapdge
 2951 dqvalratgv sgrrlvrhtv ealptaaeft atgtvlitgg tggglgaevav
 3001 wlaragaqhl vltstrrgpda pgaaelrael egygpsvsvv acdvadrdal
 5 3051 aavltalpee lpltgvvhta gvghygpldt lstaefaglt aaklagaahl
 3101 dalladreld ffflfgsiag vwsgsnqsay gaanayldal alhrrargla
 3151 atsvawgpwa eagmaaddav setlrrqglg lldpapamte lrravvrqdv
 3201 tvtvadvdwq ryaplftsar psaliaglpe vralaadert eqdatgasev
 3251 vtrvralaep eqlrlldlv rtesatvlgh ssadavpegr afrdvfgfsl
 10 3301 tavelrkrlg aatglslpst mvfdyptple laqylraeil gavlevagpv
 3351 atggaddepi aiigmacrfp ggvsspeqlw dlvasgtdai sefpvnrgwq
 3401 tghlfdpdpd rpgttystqg gflheadefd ptffgispre alvmdpqql
 3451 llettwesfe ragirpetlr stltgtfvgs syqeyglgag dgteghmvtg
 3501 sspsvlsgrl syvfglegpa vtvdtacsss lvalhlacqs lrngesnlav
 15 3551 aggatimttp npfiafsrqr alakdgrcka fsddadgmtl aegvgvvolve
 3601 rlsdaqrngh pvlavlrgha inqdgasngl tapngpsqqr virqalanar
 3651 lapgdidale ahgtgtplgd pieaqalfat ygrdrdpesa lllgsvksni
 3701 ghtqsaagia svikmvmalr hselpttpha dapsshvds agtvrlltqa
 3751 rawpetgrpr raavssfgis gtnahvllaq apvadtpaee rpavapvpia
 20 3801 agvvpwvta rsaaalrgga erllahaetv gtalpaagpl diglslvsar
 3851 arfehravvv ppagtdplaa lravatdgps pvvargvadv egrtvfvfpg
 3901 qgsqwvgmgs qlldesavfa eriaecaaal aeftdswlvd vlrgvvgaps
 3951 lervdvvqpa sfavmvslaa lwrsrgvlpd avvghsqgei aaavvsgals
 4001 lrdgarvval rsqaigrala grggmmsval svdvleprlv efegrsvvaa
 25 4051 vngprsvvva gepealdalh arltaddira rriavdyash shqvedlhee
 4101 llevlaelap rtsevpffst vtgdwldtar mdagywfrnl rgrvrfadav
 4151 adllaaeyra fvevsshplv smavqeaide agvpavaagt lrrdgggtdr
 4201 flsaaevfv rgvdvdwagl fegtgasrid lptyafqheh lwavppapea
 4251 vaaadpddaa fwtavedgdv saltaalgtd edsvaavlp ltswrrarrd
 30 4301 rstvdawryr vawkplggtl phpsltgtwl lvtadgidtd dvagaletyg
 4351 aevrrlvlde ecvdravlre rlagaedvtg ivsvlaaaer tdavpgtslv
 4401 lgtaltvali qalgdaeida pvwaltrgav stgradelta pvqaqvtgig
 4451 wtaalehpqr wggtdlpaa ldaraaqrla avlsgalgsd dqlairpsgv
 4501 ftrrivraea tagrpagtw prgttlvtgg sgtlaphlar wlaqrgehl
 35 4551 vlisrrgtaa pgaaelvaal aesgteatva acditdrdav aalladlkad
 4601 grtvrtvvht aatielhtld attladfdrv lhakvtgaqv laellddeel
 4651 ddfvlyssa gmwgsgahaa yvagnaylaa laehrrangl palslswgiw

4701 addlklgrvd pqmirrsgle fmdpqalalsg lqralddnen vlavadvdwe
 4751 tyhpyvtsgr ptplfdevpe vrrltaaaeq sagtvaegef aaalralsda
 4801 eqqrtilletv rteaasvlg1 ssaedltdqr afrdvfgfsl tavglrnrla
 4851 svtgltlpst mvfdypnpaa laaylhgela garsaaagaa avptgapdad
 5 4901 dpiaivgmsc rypggvgsae dlwrialdev daisgfpadr gwdaeglydp
 4951 dpdrpgrtys vqggflrdva efdpgffgis prealsmdpq qrlilletawe
 5001 afehagidpv gqrgsrtgtf vgasyqdyas gvpnsegseg hmitgtlssv
 5051 lsgrvsylfg fegpavtldt acssslvamh lacqslrnge sslalaggvs
 5101 imstpsmsvg fsrqralaed grckayadga dgmtlaegvg lvlllerlsda
 10 5151 ranghqvlav irgsavnqdg asngltapng psqgrvirqa lansavapgd
 5201 idvleghgtg talgdpieaq allatyggdr aperplllgs vksnightqm
 5251 asgvasvikl vralqegvvp kslhidrpst hvdwssgaig lltertpwpe
 5301 tgrprraavs sfgisgtnvh tileqapade aptpadpprd glvpvllsgr
 5351 geaalraqaa rllaaveerp eahltdlahs latsraaler raaviaadrd
 15 5401 tltrglrals dgrpdpglvq gtagrgrtaf lftgqgsqrp gmgrelhdry
 5451 pvfadaldev larlddgpdr plrevlfaap dsaeaalldr tgyaqpalfa
 5501 vevalfrllt swgltpdyla ghsvgelaaa hvagvlsldd actlvaargr
 5551 lmqalpegga mvaleaaede vlpllegltd rvsvaavngp rsvvvagvee
 5601 dvllladlfa adgrrtkrlr vshafhsp1m damlddfaav argltyhppt
 20 5651 ipfvsnvsgg lataeqvrtp dywvghvraa vrfadgidwl atqgdvhtfl
 5701 elgpdgvlsa maresltdps rtallptlrg drpeepalvt avaaahahga
 5751 rvdwsgyfad hgarrttlpt yafqrerywp dttaatsaht pgsaldaefw
 5801 aaverddvaa laasldldda tvtamvpalt awrrrrgeqt eldswryrvt
 5851 wkprggatap aaltgrwlv1 vphdhqdrqd dataawaadv etalgtttvr
 25 5901 ltvtttdraa laariteaag dggpfsgvls llplatgdag hpgapaaltl
 5951 tttavqalgd agidaplwnv trgavavgra eqvtapeqaa vwglgraval
 6001 elparfggtl dlpatldgga arrlravlaa tdgedavalr psgvflrrla
 6051 hapagpdtar tafdpagtv litggtggig ghvarrlard gathllltsr
 6101 rgpaapgada lraeeelga rvtlaacdaa drdalaalla elpddaplca
 30 6151 vfhtagvved hvvdaltpen faavlraktv aahhlhelta dldlaafvlf
 6201 sstagvlga gggnyaaana hldalaehrr shgltalsva wgpwagsgmv
 6251 adaaeltdrv rrggfeplap epavrallra ienddtval adidwerfqr
 6301 afaavrplpf vadlpetgra tpatatgaat glrqqlaelp eherpaavld
 6351 llrtqvaavl ghadprtved dhafrdlgfd sltilelrna lnaatglslp
 35 6401 atlvydlptp remadfl1ae llgtlptdta atvastaspk lsasfeqgggt
 6451 pfddpiavig igcrfpggvt tpeelwqlld egrdgisrfp ddrqwdlaal
 6501 gagasdtleg gfltgvadfd arffgispre alamdpqqrl llettweale

6551 ragidpttlr gsttgfvgt ngqdyptllr rsasdvagyv atgntasvms
 6601 grlsyalgle gpavtidtac ssslvalhwa gralragecd lvvaggvsvm
 6651 aspdsfvefs tggglapdgr ckafsdadg tawsevgil vlerlsaarr
 6701 nghqvlglir gtavnqdgas ngltapngls qqrviaqala darlrpadid
 5 6751 aieahgtgtt lgdpiaral itaygrdrda erplllgtvk snightqaaa
 6801 gaagvikmlm amrhgtlprt lhvgtpsshv dwsggtvall ddarpwprt
 6851 qprragvsaf gvsgtnahvv veqapeteap aapaaepape atptvvpwv
 6901 sgrsrealqa qldrhtahta ahparsaadv grslatdrtl fphravllag
 6951 pdgvreaara aaprtpgrrta flfsgggaqh almghdlyqr fpvyadaltd
 10 7001 vlaqfdtvld vplraalfaa pgtpeaalld qtgftqpalf avevalfrla
 7051 eswrltpdfv aghsigei

SEQ ID NO. 21

15 Glu(Asp)Leu Gly Phe(Leu, Val) Asp Ser Leu

SEQ ID NO. 22

GAG/C CTG/C GGC/G T/CTG/C GAC TCC/G CTG/C

20

SEQ ID NO. 23

Val Asp Thr Ala Cys Ser Ser

25 SEQ ID NO. 24

G/CGA G/CGA G/ACA/ G/CGC C/GGT GTC G/CAC

SEQ ID NO. 25

30

GTTGGTACCCCACTCCCGGTCCGCAC

SEQ ID NO. 26

35 CCAGCCGCATGCACCACC

SEQ ID NO. 27

CCG CGT CGG ATC CGC CGA C

5 SEQ ID NO. 28

AGC CTT CGA ATT CGG CGC C

SEQ ID NO: 29

10

DNA sequence of ERD48.seq

1 gatccgccga cgacaccggc cgccgcaccg tcaccgtcca cgcccgcccc
 51 gacgacaccg ccgaccgcac ctggacgctg caccgccaccg gtgtgctcgc
 15 101 caccacgcca ccggccgccc cggcgttcga caccacggtc tggccgcccc
 151 ccgacgccga acccctcacc accgacgact gctacgcaca cttcaccacc
 201 caccgcttcg cctacggccc cgcttccag ggctgcggg ccgcctggcg
 251 cgccggcgac gtgctgtacg ccgaggtcgc cctgccggag tccgccaccg
 301 acgaagcggc cgccttcggc ctgcaccgg cgctcctgga cgccggcctg
 20 351 caccgccgcg tcttcgccga cgaccgcgac accggactcc cgttctcctg
 401 ggaaggcgtc actctgcacg cctccggcgc caccgcgcta cgcgtccggc
 451 tcgccccgaa cggccccaac ggctgttcg tcaccgccgc cgaccgggac
 501 ggcaaccccg tcgccaccgt caccgcctg ctgcgccgcc ccctggacgc
 551 cgagcagttg accatccaca gcgccctgac ccgcgacgcg ctcttcacc
 25 601 tggactggac cccggtcccg cttcccgaca ccgccaactc cgcgcgcgcg
 651 gccctcctcg gcccgacac cgccgtgctc gccgacgcc tcggcgaccc
 701 ggccgtcgca cgccacgcaa ccctcgacga cctcctggcc ggggacacca
 751 ccccgcccgc caccgtcctc gtccccctcg gcgccccact cgacggcgac
 801 accgcgcagc acgcgcacgc cctcaccgc agcgcgctga ccctcgcca
 30 851 gcagtggctc gccaccgacc gcctcgccga ctcccgctg gtcttcgtca
 901 cccacggagc cgtcgccacc gacgacgcgc cccccaccga cctggccgcc
 951 gccgcggtct ggggcctgat ccgctccgcg cagaccgaga accccggcac
 1001 cttcaccctc ctgcacctg acaccgagcc cgactcgacc accgcgctca
 1051 gccgcgccct gaccctcgac gaaccacagc tctcctccg cgccggccgc
 35 1101 gccgcgcgcg ccgcctcac ccgcaccccc gccccacca ccaccacca
 1151 caccgcgtgg tcgcggacg gaacggtgtt ggtgacgggt ggtacgggtg
 1201 gtctgggtgg gttggtggcc cggcatctgg tgcggtcgtg tggggtgcgg

1251 catttggtgtg tgaccagtcg ttctgggtgtg ggtgctgcgg gtgcggccgg
1301 gttggtcgcg gagttggagt cgttggggcg gcgggttggt gttgcggcgt
1351 gtgatgtggg tgatggctcg gctgttgccg agttggttgc cgggtgtgtc
1401 gagtcgtatc cgttgtctgc ggtgggtgcat gcggctgggt tgttggtga
5 1451 cgggtgtggtg ggttcggtga cgccggagcg gttggctgcg gtgttgctgc
1501 cgaagggtga tgggtgcgtg aacctgcagt aggcgacgcg tgggtctggat
1551 ctggacgcgt ttgttgtctt ctctgtctgt gcgggtgtgt tcgggggtgc
1601 gggtcaggcc aactatgcgg cgggtaatgc gtttttgac gcgttgatgg
1651 ttcacgggtt ggctgggtgg ttgcctgggt tgcgttggtc gtgggggtgt
10 1701 tgggatcagg gtgtggggat gacggcgggg ctgacggagc gggatgtccg
1751 tcgtgctgct gagtcgggta tgccgttggt gacggttgat cagggtgtgg
1801 cgttgttcga tgcggcgttg gcgacgggga gtgcgcgtt ggtgccggtc
1851 cgtctggacc tggccgact gcgcacccgg ggcgacatcg caccgtctct
1901 ccgcggcctg gtcaaggcgc ccatccgccg cgcagccgcc accacacccg
15 1951 gcgacaccgg actgcggag cagctcacc ggctccagcg cgccgagcga
2001 cgggacaccc tcctcgcgt cgtccgcgac caggccgcga tggctctcgg
2051 ccacacctcg ggcgacggcg tcgaccgtc ccgcgccttc cgcgacctcg
2101 gcttcgactc gtcaccgcg gtcgaactcc gcaaccgcat cggcgcgcc
2151 accggcctgc ggctaccggc caccggcgtc ttcgactacc ccaccgccga
20 2201 tgccctcgcc gcacacctgc tcaccgaact gtcgggccc gacgccgagt
2251 cggaccccg cagacccggc gacccaccc cgggaccgac cgacgacccc
2301 atcgtcatca tcggcatgag ctgcccgttc cccggcgaca tcggctcgcc
2351 ggaggacctg tggcgctgc tcggcgacgg cgcgacgtc gtcaccgact
2401 tcccgaccaa ccgcggctgg gacctggaca acctctacga ccccgacccc
25 2451 gcgcagccg gcacctgta cgcccgacc ggcggtttcc tgcacgacgc
2501 cgccgacttc gacccgact tcttcggcat gagccccgc gaggccatgg
2551 ccacggactc ccagcagcgc ctgctgttgg agtcctcgtg ggaggcgatc
2601 gagcgggccg gcatcgaccc gctgacctg cgcgacagcc gcaccggcgt
2651 cttcgccggc gtcattgata gcggctacgg caccgcctc gacggcgccg
30

SEQ ID NO: 30

Translation product of SEQ ID NO. 29 (ERD48.seq)

35 1 saddtgrrtv tvharpddta drtwlhatg vlattpaaa afdtvwp
51 daeplttddc yahftthrfa ygpafqglra awragdvlya evalpesatd
101 eaaafglhpa lldaglhhaal laddrdtgpl fswegvtlha sgatalrvrl

151 apngpnnglsv taadpagnpv atvtrllarp ldaeqltihs altrdalfhl
 201 dwtpvplpdt ansappallg pdtavladal gdpavarhat lddllagdtc
 251 ppatvlvplg apldgdtaqh ahaltrsalt lvqqwlatdr ladsrlvfv
 301 hgavatddap ptdlaaaavw glirsaqten pgtftlldld tepdsttals
 5 351 raltldepql llragraraa rlttrtpaptt tthtpwsadg tvlvtgggtg
 401 lgglvarhlv rscgvrhl11 tsrsgvgaag aaglvaeles lgarvvvaac
 451 dvvgdgsavae lvagvsesyp lsavvhaagv lddgvvgslt perlaavlrp
 501 kvdgawnl1he atrgldldaf vvfssvagvf ggagqanyaa gnaflaldm
 551 hrvagglpgv slawgawdgg vgmtaglter dvrraesgm plltvdqgva
 10 601 lfdaalatgs aalvpvrld1 aaltrrgdia pllrglvkap irraaattpg
 651 dtglaeqltr lqraerrdt1 lalvrdqaam vlghtsgdgv dpsrafrdlg
 701 fdsltavelr nrigaatglr lpatavfdyp tadalaahl1 tellgpdaes
 751 dpdepgdpta gptddpivii gmscrfgpdi gspedlwrll gdgadvvtdf
 801 ptnrgwdldn lydpdpahag tsyartggf1 hdaadfdadf fgmspreama
 15 851 tdsqqrllle ssweaierag idpltlrdsr tgvfagvmys gygtrldga

SEQ ID NO. 31

CGCCGCATGCTGTTCTACCCCACGT

20

SEQ ID NO. 32

GGCGCGACCCGGTTCGGCCT

25

SEQ ID NO. 33

GCGAGCGGCCGCTTCACCCCGCAACTCA

SEQ ID NO. 34

30

CGCGAAGCTTGGCCGACTGCTCGACGTC

SEQ ID NO. 35

35

DNA sequence - 125401 bps (entire cluster):

1 gaattcggat cgatgctgct cgtgggtgtgt gccgcgggag gtgccgacgg

51 tgcggtgcgt gaatcggggc gtggtcgccg tcgcatcgcc ggtggtgtcg
101 gtgctgctcg tgctgaccgt ggtggtggcg acggcctact tgcggtgagc
151 ccggtccttg cgccgcgacc tgggcagcgg atgcacctct tcggtcgacg
201 ccaaggggcg gtggcgaggc tcgagatagg cgcggacgga gacgcccggc
5 251 tgccccggcg tctgccaggc cgcggagggg ccgctggaac gccacagacg
301 gccagctcg gccgcggcaa ggggtggcgt ggcttcgtcg gcggcggtga
351 cgtcgatgac gaccatgccg gggttcgtga catgttcctc gtggatatgc
401 acagttctgt atgacctatt tcgccccgtg gcgtaaggag tagccggcag
451 gtttcattcg aagggtggggc gcgggagcgg cgagttgacc gtgaggtgcg
10 501 tgcggtgttc ccggtcgccg cgggggggtg gggccggcag gtggaccgcc
551 gtggtcagca gggcgggggc gctgtgccag cggccgggga agaccgtgac
601 ccggcgaccg cccagcaggc ggcccgggac caggagtcgg gcggtgaagg
651 cgccctgtgt cgggtcgatg acgatctcgg cctcgagaaa gtccagttcg
701 cgctgggtca ggggatacca cgccttgaag acgctctcct tggcgctgaa
15 751 gagcagcccg tcccagtggc cgtccggacg atgggccgcc agggccacga
801 ggtgcggccg ttccgagggc agggcgatgg cgttcaggac gccggccggc
851 agcggaccgt tcggttcggc gtcgatgctc accgcggccg acagctccgc
901 gggggagacg gcggcgccgc ggtagccggc gcagtgcgtc atgctgccga
951 cgatgcccgg cggccactgc ggggcgccgc gccgattggg cagtatggcc
20 1001 cgggtccggg gggcgagccg gcgcagggcc cggcgggcga gatggcgacc
1051 ggtggtgaac tcgctgtgcc gggactccac gggccgggcg atgacctcgc
1101 gttccgagga gagcagcccg tcgcccgggc gcggccggtc gtcgtacgcc
1151 gcctcggttg cgaccgtggc gggcaggatc agttcgatca ccgcatacct
1201 ccggcgaacg gtaagaacag ggggttgctt cggggcacgg ttccgtcctt
25 1251 gacggggcgc cgtgggcggc cgggtcagcc ggccgcggcg cccgcggtgg
1301 gagtgtgggt gcgggctgcy tgcagccggg cgtacaggcc ctgcgcgcac
1351 aggagttagt cgtgggtgcc ctgctcgacg atgcggccgg cgtccatcac
1401 gacgatcagg tccgcgtcgc ggatcgtgga caggcgggtc gcgatcacga
1451 agctggtccg gccgcgccgc agggagtcca tggcgcggtg gatcaggacc
30 1501 tcggtccggg tgtccacgga gctggtggcc tcgtccagca cgaggacggc
1551 cggctctggc aggaaggccc gggccacggt cagcagctgc ttctcgccgg
1601 cgctgacggt gccgcactcg tcgtccagca cgggtgctga gccctgcggc
1651 aggggtcgga tgaagcggtc ggcacaggtc gcgcggggcc cctcctcgat
1701 gtccgcacgg caggcaccgg gtgcgccgta cgcgatgttc tccgcgatgg
35 1751 tgccgccgaa cagccagggt tcctggagca ccagcccga ggggaccgc
1801 aggtcgtcgc gggctatcgt cgcggtgtcg gtgccgtcca ggaggatgcy
1851 gccggagtcg ggttcgtaga agcgcacgag gaggttgccg aggggtgggtt

	1901	tgccggcgcc	ggtggggccg	acgatcgcca	cggtgctgcc	cggttccacg
	1951	gtcagcgaga	ggttctcgat	gaggggctg	tcgggggagt	agcggaagga
	2001	cacgtcggtg	aactcgacgc	ggccctcggc	gcgggcgggc	gtgccgggcc
	2051	ggagcgggtc	cggggcctgc	tcgggggctg	cgagcagggt	gaagacgcgc
5	2101	tgggcggagg	cgatcgccga	ctggagccgg	cccgccaccg	aggcgatctc
	2151	cacgatcggc	tggctgaact	ggcgggctga	gaggatgaac	gcctgcacgt
	2201	caccgagggt	cagggtgccg	tttatgacct	tccaggcgcc	gatgacggcc
	2251	accagcacat	agccgagggt	ggcgacgaac	atcatgaccg	gttccatggc
	2301	accggaggcg	aactgcgcct	tggccgcagc	ccggtagacc	gcgtcgttgc
10	2351	aggcgctgaa	gcgctcctcg	gccgccgcgc	gccggctgaa	gcccttgatc
	2401	agcgcatgac	cggtgcacac	ctcctccaca	tgggcgttga	gggtgccgtt
	2451	cgcgaccac	tgcgcgcgct	agtggggctg	cgcgcgcttg	ctgatccggg
	2501	ccgcgatcaa	cgccgagacc	ggcacgctga	gcagcatcac	cacggccagc
	2551	gacggcgaga	tcaccagcat	cagcaccagc	atcgtcaaca	gcgagaagat
15	2601	cgagggtgatc	agctcgcgca	gggtctgctg	gagggtctgt	tggaggttgt
	2651	cgatgtcgtt	ggtggtgcgg	ctgagcagct	caccggccgg	ctgccggtcg
	2701	aagtggcgca	gcggcagccg	ggtcagcttc	tcccgggcgt	cgcggcgcag
	2751	ttcgtggatg	gtgcgccaca	cccgggacgc	caccagccgg	ccctgcgccca
	2801	gcatgaacag	cgacgccacg	acgtagagcg	ccagcaggac	cagcagcagc
20	2851	cggccgatcg	cggcgaagtc	gatcccgggc	gccgggcccg	ggacgccgcc
	2901	gagcacgccg	tcggcgatca	gatcgggtgac	ccggccgagc	agcagcgggc
	2951	cgaacgcgtt	gagcacgatc	ccgccgacgc	ccatcgacac	ggcgagtgcc
	3001	acggagcggc	ggtgcggacg	cagcaggccg	acgagccgac	gtgccggccg
	3051	gggcgcggtc	cgctcctcct	caaggctcgt	cgcgagggcc	atgggcggct
25	3101	tcctcctcgg	tcagctgcga	gagcgcgatc	tcgcggtagg	tcgggctggt
	3151	gcgcagcagc	acgtcgtggg	tgccctgggc	gaccacgcgc	ccgcgggtcca
	3201	gcaccacgat	ccggtccgcg	tcgcggccgg	cggagatccg	ctgcgccacg
	3251	gtgatcaccg	tggcgcccgc	ggtgtacggc	accagcgcg	tccgcagcgc
	3301	cgcacgggtg	gcctggtcga	gcgccgagaa	acagtcgtcg	aagagataga
30	3351	tctccggccg	gcgcagcaac	gcccgggcga	gggacaggcg	ttggcgctgg
	3401	ccgccggaga	cattgccgcc	gccctgggtg	atctccgcgt	cgaggccgtc
	3451	cggcatccgc	gccacgaagt	cggccgcctg	ggcgacccgc	agcgctccc
	3501	acagctcctc	gtcgggtggc	tccggccgcc	cgaagcgag	attgctcgcc
	3551	acggtgccgg	agaacaggta	cggccgctgc	ggcacgaacc	cgacggcggc
35	3601	ggcgagcgtg	gcccggtcga	gctcgcgac	gtcggtgccg	ccgacccgca
	3651	ccgcgccctc	ggtggcgctg	gccagccgca	gcaccaagtt	caacagggtc
	3701	gtcttgccgc	tgccggtgct	gccgagcacg	gcgatccgct	cgcccggctc

3751 gacggtcagg tcgacgtccc gcagcacggg ctccctcggcg cccgggtagc
3801 ggtacccggc cgcgcacagt tcgatccggc cggcggggccc gcgcaccggc
3851 tgcggcgcg cggcgggcg cagctcgac cgggtgtcca ggacctccgc
3901 gatccggccg gcacagaccc gggcccgcg caccgacagg aacacgaagg
5 3951 cgagcatcac gacggacatc aggatcagcg agagatagct caggagggcg
4001 ctgagcgagc cgatcggcat ccggcccgcg tcgatccggg gggagccggg
4051 ccacagcagg gctacggtga aaccgttcac cagcagcagc acgaccggca
4101 gcatcgtcgc gatcagccga cccaccgcc gcgacaccac gaggaacggc
4151 tcgttgggtc ggcgaaccg cgcgcgtcg tggtcgtcg ggacgaagga
10 4201 ccggaccacc cgcaccccg tgatcgctc gcgcagcagc cggcccagcc
4251 ggtccagggt cagctgcac cgcgcgtaca ggggtgccat ccgggccagc
4301 agcaggccga agcagaccgc caccaccagc accagcgcca ccagcagcag
4351 tgccagcggg acgtcctggc gcagcgccag cagcacgctg cccaggcaca
4401 tcagcgggc gcagacgac atgccgaagc cgggtctggg gaggttctgc
15 4451 acctgctgca cgtcgttcac cgaccgggtc agcagggagg ggggtgccga
4501 ccggccgatc tcgcgggcg agaagtccag gatgcggcg aagagcgcg
4551 accgcagatc gcggcccatc gccgtcgcg tcggggcggc cagcgcggcc
4601 gcaccgagcg cggcgcgat ctgcaccagc gccaccacgc ccatcaccac
4651 acccagctcg gtgatgcgc cgcgtcgcc gcgcaccacc ccctggtcga
20 4701 taagtgcggc gccagcgctc ggagcagca aagtgccag gatctggagc
4751 agttgaagg cgcagagagc ggcggtgcc caggcgtagg ggcgcagctg
4801 tgccgcaga agtctcaaca gcagggagg acaccccggt tgacggcggc
4851 gtgcggccgc gcggtggagc gagtcggcg gcccgccgac ccggtcattg
4901 caccgcagtc cgctccgaaa ttctactagt gttgggggtg gcaacggact
25 4951 tttgacggc cgggtactcg gaattcccta gaaagcccg gatgcgttga
5001 cagcatcttc cgcggttg gcgctgcg gtgtctatcg tccggactgc
5051 gtattcgac ccaggctcg gccgagctc agcgtttgga cgggtctctt
5101 cgtgttcgag aagggtttgc catgtccaaa cgagcgctga tcaccggaat
5151 caccggccag gacggctcct atctcgcgga gcacctgctg tcccagggct
30 5201 accagggtg ggggtctgat cgcggccagg ccaatcccc caagtcccg
5251 gtcagccgcc tcgctccga actcgacttc atcgacggg acctgatgga
5301 ccagggcagc ctggtctccg ccgtcgacac cgtgcagccc gacgaggtct
5351 acaacctcg cgccatctc ttcgtgccga tgcctggca gcaggccgag
5401 ctggtcaccg aggtcaacgg catggcgctg ctgcgcatgc tggaagccat
35 5451 ccgcatggtc agcggactgt ccacctccc cagggtcagc ccgcgcggc
5501 agatccgctt ctaccaggcg tccagctcg agatgttcg caaggccgcc
5551 gagacgccgc agcgcgagac caccctcttc caccgcgca gccctacgg

5601 cgcggcaaaag gcgtacgggc actacatcac ccgcaactac cgcgagtcct
5651 tcggcatgta cgcggtctcc ggcgtgctct tcaaccacga atccccgcgc
5701 cgcggccagg aattcgtcac ccgcaagatc agcctggcgg tcgcccgcac
5751 caagcaaggc ctccaggaca agctggcact cggcaacctc gacgcggtgc
5 5801 gcgactgggg ctatgccggc gactacgtcc gcgccatgca cctgatgttc
5851 cagcaggacg ccggcgacga ctacgtcatc ggcaccgggc agatgacttc
5901 ggtgcgcgac gcggttcgga tcgcttcga acacgtcggc ctgaactggg
5951 aggactacgt cgtcatcgac ccgcacctgg tgcggcccg caggtcgag
6001 gtgctgtgcg ccgacagcgc caaggcccag gaccgcctcg gctggaagcc
10 6051 ggacgtcgac ttccccaccc tcatgcgcac gatggtcgat tccgacctgg
6101 cgcaggtttc ccgcgaaac caatacggcg acgtgctgct cgccgcaaac
6151 tggtagcagt tctcaagctt tcgaaaacta gtgaattcct gccggaattc
6201 cgacgacact cgccacatgg attcccccg tgagtggcga atccaggtgg
6251 cgaatccgaa cgtaccgccg aaacggcggtg gagaagtcgg acgccattca
15 6301 cgtgcggccg tcggctgtcg agatgggttg agttgagatg gacaacgaac
6351 aaaaactccg ggattacctc aagcttgca cggccgacct tcgacgcacc
6401 cggcggcgcg tccacaagct ggagtcggcg gcccggaac cgggtggccat
6451 catcggcacg acctgtcgct acccggcgcg cgtccgcagc cccgaagacc
6501 tctggcgcat ggtcgaggcc ggcgagcacg gcgtcaccac gttccccacc
20 6551 gaccgcggtt gggacctgga ggcgctggcc gccgcgccga ccgcctccgg
6601 cggattcctg cacgacgcac ccgacttcga cgcggacttc ttcggcatct
6651 cgccgcgcga ggcggtcgcc atggaccgcg aacagcgcg cgtcctggaa
6701 tccgcctggg aggcgttcga acgcgccggc atcgaccgga cgtccgtgaa
6751 gggcagccgc accggagtct tcatcggcgc gatggccag gactaccggg
25 6801 tcggccccgc cgacggcgcc gagggcttcc aactcaccgg caacaccggc
6851 agcgtgctgt ccggccgcac ctctacacc ttcggcacgg tcggccccgc
6901 cgtcaccgtc gacaccgcct gtcctctct cctcgtcgcc gtccacctcg
6951 ccacccaggc gctgcgggcc ggcgagtga ccctcgccct cgccggcggc
7001 gtcaccatca tgtccggccc cggcaccttc atcgaaatgg gccgccaggg
30 7051 cgggctctcc gccgacggcc gctgcgctc cttcggcgac accgccgacg
7101 gcaccggctg ggccgaaggc gtcggcatcc tcgtcctgga acggctgtcc
7151 gacgcgctcc gcaacggcca cgagatctc gccgtcgtcc gcggcaccgc
7201 cgtcaaccag gacggcgcc ccaacggcct gaccgcccc aacggcccct
7251 cccagcagca ggtcatccag caggccctgg tcaacgccc actcgccgcc
35 7301 ggggacatcg acgtcgtcga ggcgcacggc accggcacca ccctcggcga
7351 ccccgctgag gccaggccc tgctcgccac ctacgggcag aaccgcccgg
7401 cggaccggcc gctgctgtg ggctcggta agtccaacct cagccacacc

7451 caggccgccc cggcgctcgc cggcgtgatc aagatggtca tggcgatgcg
7501 gcacggcacc ctgccgcgca ccctgcacgc cgaggagccc acccaccacg
7551 tcgactggtc gcagggcgcc gtgcggctgc tgaccgacac caccgactgg
7601 cccgccaccg gggcgccgcg ccgcgccgcc gtctcctcct tcggcatcag
5 7651 cggcaccaac gcccacacca tcatcgagca ggccccgaa ccgcagcccc
7701 aggcgcgcgc gaccgcgcag gacgacgccg ccggcagcac gccggccacc
7751 gccccgtag tgcccggcgt cgtaccggtc ctgctctccg gccgcacccc
7801 ggacgccctg cgcggccagg ccgcggccct gcgcgcgcc ctcgacaccg
7851 gccggcggcc cgacctgctc gacctcgac actcctctgc caccaccgcg
10 7901 gccgggttcg agcaccgcgc cgtcctctc gccaccgacc accccgccct
7951 gaccgacggc ctcaccgccc tcgccgacgc cgacgaccg gccgccgcc
8001 ccgcctggat caccggcacc acccggggcc agaccggct cgcctctg
8051 ttcaccggcc agggcgccca acgcctcggc gcgggacggg aactcgccgc
8101 ccgtttcccg gcgttcgccca ccgccctcga cgcggcgctc gacgccttca
15 8151 ccccgacact cgaccgcccc ctgcgcgagg tcctgtgggg caccgacgcc
8201 gccctgctcg accgcaccgc atacgcccag ccggccctct tcgccgtcga
8251 agtggcgctc taccggctga tcgaatcgtt cggcgtccgc cccgaccacc
8301 tcgcggcca ctccgtcggc gagatcgtcg ccgcgcacct cgcggggtc
8351 ctctccctgg ccgacgccgc caccctcgtc gccgcccgcg gtcgcctgat
20 8401 gcaggcgctg cccgacggcg gggcgatgat cgccgtccag gcgtcggaag
8451 ccgacgtcgc cccgtgctc gccgggcacg aggaccaggt cgcgatcgcc
8501 gccgtcaacg gcccctccgc cgtcgtcctg tcggcgccg aagccaccgt
8551 caccgcgctc gccgaacagc tcgccgccga cggccgcaag acccgccggc
8601 tgcgcgctc gcacgccttc cactcgccgc tcattggagcc gatgctcgac
25 8651 gccttcgcg ccgtcgtcga agacctcacg ctccagccgc cgctcctgcc
8701 ggtcgtctcc aacctgaccg gcaagcccgc caccgtcgcc caactcacct
8751 ccgccgacta ctgggtcgac cacgtccggc acgcgtccg cttcgccgac
8801 ggcacgact ggctcgcccg gcacgacacc accgccttcc tcgaactcgg
8851 cccgcacggc gtgctgtccg ccatggccca ggactgcctg gacgcccgccg
30 8901 acgcagacgc cgtcacctc ccgccctgc gcgcggggcg ccccgaggag
8951 cacaccctca ccaccgccct cgcgggtctg caggtccacg gcgccaccct
9001 ggactggacc ggctgcttcg ccggcaccgg cgcgcccgcc accgacctgc
9051 cgacctacgc cttccagcgc cgcgctact ggcccaaggc cctccagagc
9101 ggcaccgccg acctgcgctc ggtcggcctc ggtgccgccc accaccgct
35 9151 gctctccgcc gccgtctccc tcgccgacgc aggcggcacc ctgctcaccg
9201 gccgcctctc ccggcagacc caccctggc tcgccgacca caccgtccgc
9251 ggcaccaccc tgctgcccgg taccgccttc ctgaactcg ccgtccgcgc

9301 cggcgacgag gtcggctgcg accgcgtcga ggaactcacc ctgcgcgcac
9351 cgctcctgct gcccgaacag ggcggcgtcc aggtccagtt gtggatcggc
9401 aaccccgacg tgtccggtcg ccgcaccgtc aacgtccacg cccgccccga
9451 caccggcgac gacaccccct ggaccgcca cgccaccggc gtcctcacca
5 9501 ccgccgacgc ctcccgccag ctcccggctt cgtccgagca gggcggcacc
9551 cccctcgccg gcgaccccca ccccgccctc gacgcggccc agtggccccc
9601 ggccggcgcc gaaccgctgc cgctggacgg ccactacgac cgctcgccg
9651 acggcggtt cggtacggc ccggtcttcc agggcctgcg cgccgcctgg
9701 cgcgggcgcg acgtcgtcta cgccgaggtc gagctgcccc aggcggcgccg
10 9751 gtccgacgcc gaggcgttcg gcctccacc cgccctgctc gacgcgccc
9801 tgcacgccgc gcccttcacc ggctcgggc aacgcggccg gggcggcctg
9851 ccgttctcct gggagggcgt ctccctccac gccggcgggc ccaccacct
9901 ccgcgtccgc ctgaccccg tcgccgacga cgcgctcgcc ctgaccgtcg
9951 ccgacggcac cggcgcgccc gtgctgtccg tcgactcgct cgtcctgcgc
15 10001 agcgtggcga cccaacagct cgacacggc gccgcccgtc cccgtgacgc
10051 cctcttcgc ctgactgga ccccgctcca gccgaccgc accgaccccg
10101 ggcccgtcgc cctcctcggc gccgaccct tcggcctgct caccacgcc
10151 ggattcgccg acgccccggc ataccggac ctgcgcgcc tcgccgggc
10201 ggacggcccg gtcccacca ccgtcgtgct gtccctcgcc ggacccggg
20 10251 acgacgcggc cgaccggcc cggtccgcac accgtgccc cgcgaggcc
10301 ctgcgcccg tacagacctg gctcgaccac catgagcgt tcgcccgcc
10351 ccgcctggtc ttcgtgacc gtggtgcgac ggtcgggcgt gatgttgctg
10401 ctgctgcggt gtgggtctg gtgcgttcg cgagtcgga gaatccggg
10451 tgttttgctc tggtcgatct ggatccggat ggtgcggtgg gtgcggtgc
25 10501 gtcgctcgt gcgttggtca gtggtgagc gcagcttgc gtgcgcggtg
10551 atgtgttgcg ggtcgcgct ctggtgcggc ggccgctcac cgaggtcgtt
10601 gcgggtgctg atggcaccg ggatggcgtc ggggatggct ctggtgtgtc
10651 gttctcgggt gaggtgcgg tcctggtcac tgggtgtacg ggtggtctgg
10701 gtgcggtgtt ggcgcgtcat ctggtggccg agtatgggt gcgggatctg
30 10751 ctgttggtca gtcgcagtgg tgaacgtgcc gtgggtgctg gggagtgtgt
10801 ggccgagctt gcgggtgtgg gtgcgcgggt gcgggtggt gcgtgtgatg
10851 tgaccgatcg tgcccggtg gtggagttg ttggcgggca tgcggtgtcc
10901 gcggtggttc atgcggctg tgtgctggat gacggcatgg tgggtgcgtt
10951 gaccggggag cggttgtccg cgggtgctgc gccgaagggt gatgtgtct
35 11001 ggcattctaca tgaggcgacc cgcggcctgg acctggacgc gttcgtcgtc
11051 ttctcctccc tcgccgggt cttcggcagt cccggccagg ccaactacgc
11101 ggccgcgaac gccttctcgt acgcgctgat gacgcggcgc cggcgagg

11151 gactgcccgg cctgtcactc gcatggggac cgtgggagca gtcgggcgga
11201 atgacgggca ccctgacgga cgtcgacgcc gaacggctgg cccgctccgg
11251 tgtcccgcg ctctccgtgg cgcagggcct ggccctcttc gacgctgccg
11301 tggccgggac cgacgccacc tgcgttcggg tccgcctgga cctccccgtc
5 11351 ctccgcgcac ggggtgaagt gccgccgtg ctgaggctgt tgatccgct
11401 ccgggcgcgc cgagccgccg tcgccggctc cgccaccgcg ggcaacctcg
11451 cccagcgcct gcgccgcctg gacgaggacg gccgcgacga gatggtcctg
11501 gacctcgtcc gcggtcaggt cgccctcgtc ctcggccacg cgaccggtgg
11551 cgacgtcgac gccggccgtg ccttcgcga cctcggcttc gactcgtga
10 11601 ccgccgtcga actgcgaac cgcctcaaca ccgtcaccgg cctgcccctg
11651 cccgccaccc tggctcttga ctaccgacc gtccggcacc tcgccacgta
11701 cgtcctggac gagttgttgg gcacggatgc cgaggtggcg accgtgcagc
11751 cgccgcggt tgcggtggcg gacgatccga tcgtcatcgt gggcatggcc
11801 tgccgtacc ccggtggcgt cagctcccc gaggacctgt ggcgctgct
15 11851 caccgaaggc accgacgccg tctcgggctt cccgaccaac cgtggttggg
11901 acgtcgaatc cctctatcac ccggaccctg accaccggg tacctctac
11951 acgcgtcgg gtgggttctt gcatgaggcg ggggagttcg atccggggtt
12001 cttcgggatg agtcgcggg aggcgttggc gaccgattcc cagcagcgt
12051 tgttgttga gtcgtcgtgg gaggcgatcg agcgggccg tattgatccg
20 12101 gtgagtttgc ggggtagtcg gacgggtgtg ttcgcggggg tgatgtacag
12151 cgattacagc gcgatgttgg cgagtcgga gttcgagggt ttccagggca
12201 gtgggagttc gccgagttg gcgtcgggtc ggggtggccta cacgttggg
12251 ttggaaggcc cggcggtgac ggtggatacg gcgtgttcgt cgtcgttgg
12301 ggcgatgcac tgggcgatgc aggcgttgcg tagtggtgag tgtgggttgg
25 12351 cgttggccgg tgggtgtgac gtgatgtcga cgctgcggt gtttgtggac
12401 tttgctcggc agcggggtt gtcgccgat ggccggtgca aggcgttgc
12451 ggatgcggcc gatggtgtgg gctggtccga gggcgtcggc gtgttggcc
12501 tggagcggca gtcggacgcg gtgcgcaatg gtcacgagat tttggctgtg
12551 gtgcggggtt cggcggtcaa ccaggatggt gcgtccaatg gtttgacggc
30 12601 gccaatggt ccgtcgcagc agcgggtgat ccggcaggcg ttggccagt
12651 gtggtctgac ggccggtgac gtggatgtgg tggaggcgca tggtaggg
12701 acgacgctcg gtgatccgat cgaggcgag gcgttgttgg cgacgtatgg
12751 gcgggatcgt gagcctgagc ggccgttgtt gttgggttcg gtgaagtcga
12801 atctggggca tacgcaggct gctgcgggtg tggcgggcgt tatcaagatg
35 12851 gtgttggcga tgcggcatgg tgtggtgccg cggacgttgc atgtggatgc
12901 gccttcttcg catgtggact ggtccgaggg tgcggtggag ctgctcagt
12951 agcaggcggc ctggccggag acgggtcggg tgcggcgggc ggggtgtctc

	13001	tccttcggca	tcagcggtag	caatgtgcat	gtcatcgtcg	agcaggcgcc
	13051	gggcgccaag	gcgatcgccg	cggccggtgc	ggcgcggcgc	acgccgggtg
	13101	ccgtgccggt	gctgctctcg	gggcgtggcc	ggagtgcctt	gcggggccag
	13151	gccgcccgcc	tgctcggaca	cctccaggcc	cgacccgacg	ccgaactcgt
5	13201	cgatgtcgca	ctgtcgttgg	cgaccacccg	gtcccgttcc	gagcagcggg
	13251	ccgcgcgtcg	ggcgcaggac	cgcgaccagc	tgatcgccct	gctggggggcg
	13301	ctggccgccc	accgccccga	ccccgcccgc	gtcgagggcg	aggccgcccg
	13351	acgcggcccg	accgcggtgc	tggtcaccgg	acagggcagc	cagcggggcg
	13401	ccatggggcg	tgaactccac	gaggtgcagc	cggagtccgc	cgcggcgttc
10	13451	gacgcggtgt	gtgccgtttt	cgacccgctg	ttggaccggc	cgctgcgcga
	13501	ggtggtgttc	gccgaggacg	gcagcgacga	ggccgcactg	ctggacgaga
	13551	ccggttgga	gcagccggct	ctgttcgccc	tcgaggtggc	gctgttccgg
	13601	ctggtggaga	gttgggggtg	ccgtccggac	ttcgtggccg	gccattccat
	13651	cggtgagatc	gcggcgccgc	acgtcgccgg	ggtgctgacg	ttggaggacg
15	13701	cctgccgtct	ggtggccgcg	cgggcgacgc	tgatgcaggc	gctgccgacc
	13751	ggcggcgcg	tgatcgcgat	ccaggccacc	gaggacgaga	tcgcggcgca
	13801	cctcgacgac	acggtggcga	tcgccgccgt	caacggggccg	cagtccgtgg
	13851	tgatctccgg	tgacgaggag	gccgccgaaa	cgatcgccgc	cacgttcgcc
	13901	gaacgcgggc	gcaagaccaa	gcggctgcgg	gtgagccatg	ccttccactc
20	13951	gccgcggatg	gacgggatgc	tggacgcctt	ccggatcgtc	gccgaggggc
	14001	tgacctaccg	ggcgcgcgcg	atcccgtctg	tctccgacct	caccggcccg
	14051	cgcgccgacg	atgcggaggt	gtgcaccgcg	gagtactggg	tgccggcacgt
	14101	ccgagaggcc	gtgcggttcg	ccgactgcgt	gcggacgctg	cgcgacgccg
	14151	gggccaccac	cttcttgga	ctgggctccg	acggcctgct	gaccgcgatg
25	14201	gccgaggaca	ccctcgggtg	cgaccacgac	gccgaactgg	tgccgatgct
	14251	gcgcgcgggg	cgcgccgagg	aactggccgc	ggccaccgcc	ctggcccggc
	14301	tccaggtgcg	cggcgtggac	gtggactggg	cggcgtacct	cgccggcacc
	14351	ggcgcgccg	gcaccgacct	gccgacctac	gccttcacgc	acgcgtacta
	14401	ctggccgcag	ctgccgacct	cggccgcgcg	cctcgccgcc	gccgatcccg
30	14451	ccgaccagca	gctgtggggc	gctgtggagc	gcggcgacgc	ccgcgaactc
	14501	gccgacatcc	tcggcctggg	cgaacaggac	ctcacgccgc	tggaactcct
	14551	gctgcccggc	ctcacctcgt	ggcggcgccg	caaccaggag	aagcacctcc
	14601	tggaaccctt	gcgtaccgcg	gtggagtggg	cacgactgag	caagccgacc
	14651	gccccgggtc	tcgacggcac	ctggctgctg	gtcgccctcc	acgccaccgc
35	14701	ggccgaccag	ccagccctcc	tcgacggcct	ggccgacgcc	ctcggtctcg
	14751	acggcgcgcg	ggtgcgtcgc	ctgcttctgg	acgactcctg	cgcggaaccg
	14801	gcggtgctcg	ccgaacgact	ggcgcggacc	gccgacgtgg	acgccgcgac

14851 ccaggtgctg tccgtgctgc cgctcgacga gcgggacgcc gacgactgcc
14901 cgccgctcac ccgcggaactg gcgctgaccg tcgcgctcgt ccaggccctc
14951 gccgacaccg gcgcccaggg ccggctgtgg accgccaccc gcggcgccgt
15001 ctccaccaac ccgcccagcc cggtcaccca ccccgctccag gccgctgcct
5 15051 ggggcctggg ccggggcgtc gccctggagc acccacggct gtggggcggc
15101 ctggtcgacc tgccgcaggt cttcgacgag cgggccggac agcggtcgc
15151 cgggatectc gccgtcaagg acgcaccgga cggcgaggac cagggtggcg
15201 tgcgggccac cggagtctcc ggccgcggc tcgtccgcca caccgtcgaa
15251 gcgctgcca cgcccgcgga gttaccgcc accggcaact tcctgatcac
10 15301 tggtagcacc ggtggcctgg gcgccgaggt cggccggtgg ctggcccgcg
15351 ccggcgcca gcacctctc ctgaccagcc gcccgggccc ggacgcgccg
15401 ggcggcgccg aactccgggc cgaactggag ggctacgggc cgtcgggtgc
15451 cgtcgtcgcc tgcgacgtcg ccgaccggga cgcgctcgcc gccgtcctca
15501 ccgcaactgc cgaggaactg ccgctgaccg gtgtcgtgca caccgcaggc
15 15551 gtcggccact acggcccgct ggacaccctg agcaccgccg agttcgccgg
15601 cctcaccgcc gccaaactcg ccggcgccgc ccacctcgac gccctgctcg
15651 ccgaccgcca actggacttc ttcgtcctct tcggctccat cggcggtgtc
15701 tggggcagtg gcaaccagag cgcctacggc gccgccaacg cctacctgca
15751 cgcgctcgcc ctgcaccgcc gcgcccggcg cctcgccggg acctccgtcg
20 15801 cctggggccc gtgggcccag gccggcatgg ccgccgacga tggcgtttcc
15851 gagaccctgc gccgccaggg cctcggcctg ctcgaccggg ccccgcccat
15901 gaccgagctg cggcgccgg tcgtccggca ggacgtcacc gtcaccgtcg
15951 ccgacgtgga ctggcagcgc tacgcaccgc tgttcacctc cggccggccc
16001 agcgccctga tcggcgccct gcccgaggtc cgcgccctcg ccggcgacga
25 16051 gcgcaccgag caggacgcca ccggcgctc cgaggtcgtc acccgcgctc
16101 gcgcccctgg cgaaccgag caactgcgcc tgcgaccga cctcgtccgc
16151 accgagtcgg ccacctcct cggccacagc tcgcccagc cgtgcccga
16201 gggccgcgcc ttccgcgacg tcggcttcga ctcgctgacc gcggtcgagc
16251 tccgcaagcg cctggcgccc gcgaccgggc tgtccctgcc cagcaccatg
30 16301 gtcttcgact acccgacacc gctggaactc gccagtagc tgcgggaggc
16351 gatectcggc gcggtgctgg aagtcggcg cccggtcgcc accggcgggc
16401 ccgacgacga gccgatcgcc atcatcggca tggcctgccc cttccccggc
16451 ggcgtcagct ccccggaaca gctgtgggac ctggtcgcc cgggaccga
16501 cgcgatcagc gagttccccg tcaaccggg ctggcagacc gggcacctct
35 16551 tcgaccgga ccccgaccgg ccgggcacca cctactccac ccaggcgggc
16601 ttctccacg aggcgacga gttcgacccc accttcttcg gcctctcgcc
16651 ccgcgaggcg ctggtcatgg acccgacga gcggctcctg ctggagacca

	16701	cctgggagtc	cttcgagcgc	gccgggatcc	gcccggaaac	cctccgatcc
	16751	accctgaccg	gcaccttcgt	cggctccagc	taccaggagt	acggcctggg
	16801	cgcgggcgac	ggcaccgagg	gccacatggt	caccggcagc	agccccagtg
	16851	tgctctccgg	ccgactgtcg	tacgtcttcg	gtctggaagg	cccggcggtc
5	16901	accgtcgaca	ccgcctgctc	gtcctcgctc	gtggcgctgc	acctggcctg
	16951	ccagtcgctg	cgcaacggcg	agagcaacct	ggccgctgcc	ggcggcgcca
	17001	cgatcatgac	gacgccaac	ccgttcacgc	cgttcagccg	gcagcgcgcc
	17051	ctcgccaagg	acggccgctg	caaggcggtc	tccgacgacg	cggacggcat
	17101	gacgctcgcc	gagggcgctg	gcgtcgctct	cgctcgagcg	ctctccgacg
10	17151	cgcagcgcaa	cggccacccg	gtcctggcgg	tcctccgagg	ctccgccatc
	17201	aaccaggacg	gcgcctccaa	cggcctgacc	gcgccaacg	gcccgtccca
	17251	gcagaggggc	atccgccagg	ccctcgccaa	cgcccgctc	gcgcccgggg
	17301	acatcgacgc	cctggaggcg	cacggcaccg	gcacaccgct	cggcgacccc
	17351	atcgaggccc	aggcactggt	cgccacctac	ggccgagacc	gcgaccccga
15	17401	gagcgcgctg	ctgctcggtc	cgggtgaagt	caacatcggc	cacaccaggt
	17451	ccgccgcggg	catcgccagc	gtcatcaaga	tggatcatgg	gctgcgccac
	17501	tccgaactgc	cgccgacct	gcacgcccgc	gcgcgctcct	cgcacgtgga
	17551	ctggtcggcc	gggacgggtc	ggctgctgac	ccaggcgcg	gcctggccgg
	17601	agaccggctg	cccgcgccgg	gccgcggtgt	cctcgcttcg	catcagcggt
20	17651	accaacgccc	atgtcctgct	ggagcaggcg	cccgtcgagg	acacccgggc
	17701	cgaggagcgg	cccgcctggg	cgccgggtcc	gacgcgccgc	ggcgctgctc
	17751	cgtgggtggt	caccgcccgc	agcgccgccc	ccctgcgcgg	ccaggccgag
	17801	cgctcctcgc	cgacgcccga	aaccgtcgga	accgcccctg	cggccgcccg
	17851	accgctcgac	atcggcctgt	cgctgggtct	cgcgcgcgcc	cgtttcgagc
25	17901	accgtgccgt	cgctgtcccg	cccgcgggca	ccgacccgct	ggccgcgctg
	17951	cgcgcgctgc	cgacggagcg	gccctcgccc	gtggtcgccc	gtggcggtgc
	18001	ggacgctgag	ggtcggagcg	tgctcggtgt	ccccggtcag	ggttcgcagt
	18051	gggtggggat	gggttcccaa	ctccttgatg	agtcggcggt	gttcgcggag
	18101	cggattgccg	agtgtgcggc	ggcactcgcc	gagttcaccg	actggtcgct
30	18151	ggtcgatgtg	ctgcgggggtg	tggtgggtgc	gccgtcggtg	gagcggtcgc
	18201	atgtggtgca	gccggcgctg	ttcgcggtga	tggtgtcggt	ggctgcgttg
	18251	tgcggttccc	gtggtgtggt	gccggtatgc	gtggtggggc	attcgagggg
	18301	tgagatcgct	gctgcgggtg	tgctgggtgc	gctgtcggtg	cgggacgggg
	18351	cgcgggtggt	ggcgctgcgg	agtcaggcca	ttggtcggtg	gttggcgggg
35	18401	cggggaggga	tgatgtccgt	cgcgctgctg	gtggacgtgc	tcgaaccgcg
	18451	gttggctgag	ttcgaggggc	gggtgtcggt	ggccgcccgc	aacggcccgc
	18501	gctccgtcgt	ggtcgcccgc	gagcccaggg	cgctggacgc	gctgcacgcc

18551 cggtgaccg ccgacgacat cggggcccg cggatcgcg tggactacgc
18601 ctgcactcg caccaggtcg aggacctgca cgaggaactg ctggaggtgc
18651 tggcggagct ggccgcccgc acgtcggagg tgcggttctt ctcgaccgtg
18701 accggcgact ggctggacac cgcgcggatg gacgccggct actggttccg
5 18751 caacctgcgc ggacgggtgc ggttcgcgga cgcggtggcg gacctgctgg
18801 cggcggagta ccgcgcattc gtcgaggta gctcgaccc ggtgctgtcg
18851 atggcgggtgc aggaggcgat cgacgaggcc ggcgtgccgg ccgtcgccgc
18901 cggcaccctg cgcgcgcacc agggcggcac cgaccgcttc ctgctgtcgg
18951 ccgcggaggt cttcgtgcgc ggtgtggacg tggactgggc ggggctgttc
10 19001 gaggggaccg gtgcgtcccg gatcgacctg ccacactacg ccttccaaca
19051 cgaacacctg tgggcccgtc cgcgcgcgc ggaggccgtc gccgcgcgcg
19101 acccggacga cgcggccttc tggaccgagg tcgaggacgg tgacgtctcc
19151 gcgtgaccg ccgcgtcgg caccgacgag gactccgtcg ccgcgtgct
19201 gccgcctcg acctcctggc gccgggccc cgcgcgcgc tccaccgtgg
15 19251 acgcctggcg ctaccgcgtc gcctggaaac ccctcggcg caccctgccg
19301 caccgctccc tgaccggcac ctggctgctg gtcaccgccc acggcatcga
19351 cgacaccgat gtggcagggg cgttgagac ctacggcgcc gagggtgcgc
19401 ggctggtcct ggacgaggag tgcgtcgacc gcgcgtcct gcgggagcgg
19451 ctggccggcg cggaggacgt gaccggcatc gtctccgtcc tcgccgcgc
20 19501 cgagcggacc gacgcggtac cgggcacctc cctggtgtc ggacccgcc
19551 tgaccgtggc actgatccag gccctcggcg acgcgaaat cgacgctccc
19601 gtatgggctg tgaccgcgg cgcggtctcc accggccggg ccgacgagct
19651 gaccgcgcc gtccaggcac aggtcaccgg catcggtgg accgcggcg
19701 tggagcacc gcagcgtgg ggcggcacc tcgacctgcc cgcgcctc
25 19751 gacgcccggg ccgccagcg gtcgcgcgc gtgctgtccg gcgcctcgg
19801 cagcgacgac cagctggcca tccggccctc cggggtcttc accgcgcga
19851 tcgtgcgggc cgaggccacc gccggcggg ccgcggcac ctggacgccg
19901 cgcggcacca cactggtcac cggcggtcc ggacacctc cccgcacct
19951 cgcgcgtgg ctggcccaac gcggcgccga gcacctggtc ctgatcagcc
30 20001 ggcgcggcac ggccgcccc ggccgcgcgc aactcgtcg ggaactggcc
20051 gagtcgggca ccgaggcgac cgtcgcgcgc tgcgacatca ccgaccgga
20101 cgcggtcgcc gcgtgctgg ccgacctcaa ggccgacgg cgcacctcc
20151 gcacctcgt gcacaccgcc gccaccatcg agctgcacac cctggacgcc
20201 accacctgg cggacttcga ccgggtgctg cagccaagg tcaccggcg
35 20251 ccaggctctc gccgaactgc tcgacgacga agagctggac gacttcgtcc
20301 tgtactctc caccgcggc atgtgggga gcggcgcca cgcgcctac
20351 gtcgcggca acgctacct cgcgcgtc gccgagacc gccgggcaa

	20401	cggaactgccc gccctgtcgc tgtcctgggg catctggggc gacgacctca
	20451	aactggggccg ggatgatccc cagatgatcc ggcgcagcgg cctggagttc
	20501	atggaccgcg agctggccct gagcggcctg cagcgggcgc tggacgacaa
	20551	cgagaacgtg ctgcgggtcg ccgacgtgga ctgggagacc taccaccccg
5	20601	tctacacctc cggccgaccc accccgtctt tcgacgaggt gccggaggtc
	20651	cgccgggtca ccgcggccgc cgagcagagc gccgggaccg tcgccgaggg
	20701	cgagttcgcc gccgcgtgc gcgccctgtc cgacgccgag cagcagcgca
	20751	ccctgctgga gaccgtccgc accgaggcgg cgtccgtcct cgggctgtcc
	20801	tccgccgagg acctcaccga ccagcgggcc ttccgcgacg tcggcttcga
10	20851	ctcgctgacc gccgtcggcc tgcgcaaccg gctgcctcc gtcaccggcc
	20901	tgaagctgcc ctgcagatg gtcttcgact accccaacct ggccgcgtc
	20951	ggccctatc tgcacggcga gctggccggc gcccggtccg ccgcgcgcgg
	21001	cgccgcgcgc gtcccgaccg gcgccccga cgccgacgac ccgatcgca
	21051	tcgtcgcat gagctgccgc taccgcggc gggtcggtc cgccgaggac
15	21101	ctgtggcgga tcgccctgga cgaggtcgac gcgatctccg gcttccccgc
	21151	cgaccgcggc tgggacgccg agggcctcta cgaccggac ccgacccggc
	21201	ccggccgcac ctactccgtc caggcgcgat tcctgcgcga cgtcgccgag
	21251	ttcgaccggc gcttcttcgg gatctcgccg cgcgaggcgc tgtcgatgga
	21301	cccgcagcag cggtcctgc tggagaccgc ctgggaggcg ttcgagcacg
20	21351	ccggcatcga ccgggtcggc cagcgcggca gccgcaccgg cacttcgtc
	21401	ggcgcagct accaggacta cgcctccggc gtgcccaaca gcgagggtc
	21451	cgaaggccac atgatcaccg gcacgtctc cagtgtgctg tccggccggg
	21501	tgtcctacct cttcggcttc gagggccccg ccgtcacgt cgacaccgcc
	21551	tgtctctct ccctggtcgc catgcacctg gcctgccagt ccctgcgcaa
25	21601	cggggagagc tcgctggccc tggccggcgg cgtcagcatc atgtccacct
	21651	cgatgtcggt cgtcggttc agccggcagc gcgccctgc cgaggacggc
	21701	cgctgcaagg cgtacgcgga cggcgccgac gggatgacct tcgccgagg
	21751	cgtcggcctg gtgtgtgtg agcggctgtc cgacgcccgc gccaacgggc
	21801	accaggtgct cgccgtgatc cgcgggtccg ccgtcaacca ggacggcgcc
30	21851	tccaacggcc tgaccgcacc caacggcccg tcccagcagc gcgtcatccg
	21901	ccaggcgtg gccaaactcc cgggtggcgc cggcgacatc gacgtcctgg
	21951	agggccacgg caccggatcc gccctcggcg accccatcga ggccgaggcc
	22001	ctgtctgcca cctacggcca ggaccgcgc cccgaacggc cgctgtgtct
	22051	cggtcgggtg aagtccaaca tcggccacac ccagatggca tccggcgctc
35	22101	ccagcgtcat caagctcgtc cgcgccctcc aggaaggcgt ggtgcccaag
	22151	tccttgcaca tcgaccggcc ctccaccac gtcgactggt cctcgggcgc
	22201	catcgggctg ctaccgaac gcaccccggt gcccgagacc ggccggccgc

20401 cggactgccc gccctgtcgc tgtcctgggg catctggggc gacgacctca
20451 aactggggccg ggtcgatccc cagatgatcc ggcgcagcgg cctggagttc
20501 atggacccgc agctggccct gagcggcctg cagcggggcg tggacgacaa
20551 cgagaacgtg ctcgcggtcg ccgacgtgga ctgggagacc taccaccccg
5 20601 tctacacctc cgcccgaccc accccgctct tcgacgaggt gccggaggtc
20651 cgccggctca ccgcgccgc cgagcagagc gccgggaccg tcgcccaggg
20701 cgagtctgcc gccgcgtgc gcgccctgtc cgacgccgag cagcagcgca
20751 ccctgctgga gaccgtccgc accgaggcgg cgtccgtcct cgggctgtcc
20801 tccgcccagg acctcaccga ccagcgggcc ttccgcgacg tcggcttcga
10 20851 ctcgctgacc gccgtcggcc tgcgcaaccg gctcgccctc gtcaccggcc
20901 tgacgtgcc ctcgacgatg gtcttcgact accccaacct ggccgcgtc
20951 gccgcctatc tgacggcgga gctggccggc gcccggtccg ccgcccgcgg
21001 cgccgccgcc gtcccagacc gcgccccga cgccgacgac ccgatcgga
21051 tcgtcggcat gagctgccgc taccgcggcg gggtcggctc cgccgaggac
15 21101 ctgtggcgga tcgccctgga cgaggctgac gcgatctccg gcttccccgc
21151 cgaccgcggc tgggacggcg agggcctcta cgaccggac ccgaccggc
21201 ccggccgcac ctactccgtc cagggcggat tcctgcgga cgtcgccgag
21251 ttcgacccgg gcttcttcgg gatctcgccg cgcgaggcg tgctgatgga
21301 cccgcagcag cggctcctgc tggagaccgc ctgggaggcg ttcgagcacg
20 21351 ccggcatcga cccggtcggc cagcgcggca gccgcaccgg caccttcgtc
21401 ggccgagctc accaggacta cgcctccggc gtgccaaca gcgagggtc
21451 cgaagccac atgatcaccg gcacgtctc cagtgtgtg tccggccggg
21501 tgtcctacct cttcggttc gagggccccg ccgtcacgt cgacaccgcc
21551 tgctcctcct ccctggtcgc catgcacctg gcctgccagt ccctgcgaa
25 21601 cggggagagc tcgtggccc tggccggcgg cgtcagcatc atgtccacc
21651 cgatgtcgtt cgtcggttc agccggcagc gcgccctcgc cgaggacggc
21701 cgctgcaagg cgtacgcgga cggcgccgac gggatgacct tcgccaggg
21751 cgtcggcctg gtgtgtgtg agcggctgtc cgacggccgc gccaacgggc
21801 accagtgct cgccgtgatc cgcggctccg ccgtcaacca ggacggcgcc
30 21851 tccaacggcc tgaccgcacc caacggcccg tcccagcagc gcgtcatccg
21901 ccaggcgtg gccaaactcc cggtggcgcc cggcgacatc gacgtcctgg
21951 agggccacgg caccggtacc gccctcggcg accccatcga ggcgaggcc
22001 ctgctcgcca cctacggcca ggaccgcgc cccgaacggc cgtgtgtgt
22051 cggctcgtg aagtcaca tcggccacac ccagatggca tccggcgctg
35 22101 ccagcgtcat caagctcgtc cgcgccctcc aggaaggcgt ggtgcccaag
22151 tccctgcaca tcgaccggcc ctccaccac gtcgactggt cctcggcgcc
22201 catcgggctg ctcaccgaac gcaccccggt gcccgagacc ggccggccgc

22251 gccgcgcgcg cgtctcctcc ttcggcatca gcggcaccaa cgtccacacc
22301 atcctcgaac agggcccccgc ggacgaggcg cccacgcccg ccgacccgcc
22351 gcgggacggc ctggtgccgg tctgtctctc cggccgcggc gagggcgcgc
22401 tgcgcgccca ggccgcccgc ctgctcgctt tcgtcgagga gcggcccagag
5 22451 gcccacctca ccgacctcgc ccaactccctc gccacctcgc gcgcccgcgt
22501 ggaacgcgcg gccgccgtca tcgccgccga ccgcgacacc ctgacccgcg
22551 gcctgcgcgc cctgtccgac ggccggcccg accccggcct ggtccagggc
22601 accgcgggac gcggccggac cgccttcctg ttcaccggac agggcagcca
22651 gcgccccggc atggggccgc aactccacga ccgctacccg gtgttcgccg
10 22701 acgcgctgga cgagggtgtg gcccggtctg acgacggacc ggaccggccg
22751 ctgcgcgagg tgetgttcgc cgcgccgac tccgccgagg ccgcgctcct
22801 ggaccggacc ggctacgccc agcccgcgct gtctgccgtc gaggtcgcgc
22851 tgttcgcct gctgacgtcc tggggcctga ccccgacta cctggccggc
22901 cactccgtcg gcgaactcgc cgcgcgcac gtcgccggcg tgctgtcgct
15 22951 ggacgacgcc tgcaactctg tcgccgccg cgccgggtc atgcaggcgc
23001 tgcccgaggg cggcgcgatg gtcgccctgg aggcgcgga ggacgaggtc
23051 ctgccgctcc tggaggggct caccgaccgg gtgtccgtcg ccgccgtcaa
23101 cgggccgcgg tccgtggtcg tcgccggcgt cgaggaggac gtgtcctcc
23151 tcgccgacct cttegccgcc gacgggccc ggaccaagcg gctgcgggtg
20 23201 agccatgcct tccactcgcc gctgatggac gccatgctcg acgacttcgc
23251 cgccgtcgcc cgcgggtga cctaccacc gccgacgatc ccgttcgtgt
23301 cgaacgtcag cggggcctg gccaccgcc aacagggtcc caccgccgac
23351 tactgggtcg ggcacgtcc cgcgcggtc cgttcgccg acggcatcga
23401 ctggctcgcc acccagggcg acgtccacac cttectggag ctgggcccg
25 23451 acggcgtgct cagcgccatg gcccgggaga gcctcaccga cccgtcccgc
23501 acggcactgc tgccgacct gcgcggcgac cgcccgagg aacctgcct
23551 ggtcaccgcg gtcgccgcgg cccacgcgca cggcgccgc gtcgactgga
23601 gcgggtactt cgccgaccac ggcgcgccc ggaccacgt gccgacctac
23651 gcgttccaac gcgagcggtg ctggcccgc accaccgcc ccacgagcgc
30 23701 ccacacgccc ggatccgcc tcgacgccga gttctgggcc gccgtcgagc
23751 gggacgacgt cgccgcctc gccgcctccc tggacctgga cgacgccacc
23801 gtcaccgca tgggtcccgc gtcaccgcc tggcgccggc gccgcggcga
23851 gcagaccgaa ctggactcct ggcgtaccg cgtcacctgg aagccgcgcg
23901 gcggcgccac cgcacccgcc gccctaccg gccgctggct cgtgctcgtc
35 23951 ccgcacgacc accaggaccg tcaggacgac gcgaccgcg cctgggcagc
24001 cgacgtcgag accgccctgg gcaccaccac cgtccggctg acggtcacca
24051 ccaccgaccg cgccgcgctg gccgcccgga tcaccgaagc cgccggcgac

24101 cagggcccgt tcagcgggtgt gctgtccctg ctgccgctcg ccaccggcga
 24151 cgccggccac cccggtgctc ccgccgccct caccctcacc accaccgccg
 24201 tccaggccct cggcgacgcc ggcatcgacg cgccgctgtg gaacgtcacc
 24251 cgcggggccc tggcgtcggt ccgcgccgaa caggtcaccg cgcccgaaca
 5 24301 ggccgcccgtc tggggcctgg gccgcgccgt cgccctggaa ctgccggccc
 24351 ggttcggcgg caccctcgac ctgcccgcca ccctggacgg ccaggccgcc
 24401 cgccggttgc gcgcggtgct cgcggttacc gacggcgagg acgcggtggc
 24451 cctgcggccc tccggcgtct tcctccgccg cctggcccac gcccggccg
 24501 gcccgcacac cgcccgacc gccttcgacc cggcgggccg caccgtcctg
 10 24551 atcaccggcg gcaccggcgg catcgggcgc cagctcgccc gccgcctggc
 24601 ccgcgacggc gccaccacc tgctgtcac cagccgccgc ggcccggccg
 24651 ccccgggcgc cgacgcgtc cgcgccgaac tggaggaact gggcgcccgg
 24701 gtcaccctcg ccgcctcgca cgccgccgac cgcgacgcgc tggccgcgt
 24751 cctcgccgaa ctgcccgac agcccccgt gtgcgcggtg ttccacaccg
 15 24801 ccggcgctgt cgaggaccac gtcgtggacg cgctcacacc ggagaacttc
 24851 gccgcggtgc tgcgcgcaa gaccgtcgcc gccaccacc tgcacgagct
 24901 gaccgccgac ctggacctcg ccgcttctgt gctgttctcc tccaggccg
 24951 gcgtcctcgg cgccgccgga cagggcaact acgcgcgggc caacgcccac
 25001 ctggacgccc tcgccgaaca ccgccgtcc caccggcctga ccgcgtgtc
 20 25051 cgtgccttg ggcccgctgg ccggtccgg catggtcgcc gacgcccg
 25101 aactcaccga ccgggtacgg cgcgcggtt tcgaaccgct cgccccgaa
 25151 ccggccgtgc gcgcctgct gcgcgccatc gagaacgacg acaccaccgt
 25201 cgcgctcgcc gacatcgact gggagcgctt ccagcgcgcc ttcgcccgg
 25251 tccgcccgct gccgttcgtc gccgacctcc ccgagaccgg ccggggcacc
 25 25301 cccgcgaccg ccaccggcgc cgccaccggc ctgcggcagc aactcgccga
 25351 actgccggag cagagcgcc cggcagcgg cctggacctg ctgctatccc
 25401 aggtcgccgc cgtcctcggc caccgccgac cgcgaccgt cgaggacgac
 25451 cacgccttcc gcgacctggg cttcgactcg ctgaccatcc tggaaactgcg
 25501 caacgccctc aacgcgcga ccggcctgag cctgcccgc accctggtct
 30 25551 acgacctgcc caccgcgcg gagatggcgg acttctctgt cgccgaactc
 25601 ctcggcaccc tgcccaccga caccgccgcg accgtcgcca gcacggcctc
 25651 ccccaagctc tcagcttctg tcgagcaggg cggtagcccc ttcgacgacc
 25701 cgatcgccgt catcggcac gggtgccgt tccccggcgg cgtcaccacc
 25751 ccggaggagc tctggcagct cctcgacgag ggccgcgacg gcacgagccg
 35 25801 cttccccgac gaccgcggct gggacctcgc cgcgctgggc gccggcgcc
 25851 ccgacaccct ggagggcggc ttctgaccg gcgtcgccga cttcgacgcc
 25901 cggttcttcg gcattctgcc ccgcgaggcg ctggccatgg acccccagca

25951 gcggctgctg ctggagacca cctgggagggc gctggagcgg gccggcatcg
26001 acccgaccac gctgcgcggc tccaccaccg gcgtcttcgt cggcaccaac
26051 ggccaggact acccgacgct gttgcgcgcg tccgcctcgg acgtggccgg
26101 ctacgtcgcc accggcaaca ccgccagcgt gatgtccggc cgctgtcct
5 26151 acgcgctcgg cctcgaaggc ccggccgtca ccatcgacac cgctgtctcc
26201 tcctcgctcg tcgcctcgca ctgggcccgg ccggcgctgc gcgccggcga
26251 gtgcgacctc gtggtggccg gcggcgcttc ggcatggcc agcccgact
26301 ccttcgtcga gttctccacc cagggcggcc tggcaccgga cgggcgctgc
26351 aaggcgttct ccgacgccgc cgacggcacc gcctggctcc aaggcgctcg
10 26401 catcctcgtc ctggaacgcc tctccgcgc ccgccgaac ggccaccagg
26451 tcctcggcct gatccgcggc accgccgtca accaggacgg cgcgtccaac
26501 ggctgaccg cgcaccaacg cctctcccag cagcgctca tcgccaggc
26551 actcgccgac gccgcctgc gcccgcgga catcgacgc atcgaggcg
26601 acggcaccg caccacctc ggcgaccgga tcgagggccg cgccctgatc
15 26651 accgcctacg gccgggaccg ggacgccgaa cggccgctgc tgctgggac
26701 cgtcaagtcc aacatcggg acaccaggc cggcccggt gccgccggc
26751 tcatcaagat gctgatggc atgcgccac gaccctgcc caggacgctg
26801 cacgtgggca cccgctccag ccacgtcgac tggagcggc gcaccgtcgc
26851 gtcctcgac gacgcgcggc cctggccacg gaccggcag ccgcggcgc
20 26901 ccggcgcttc cgcttcggc gtcagcgga ccaacgccc cgtcgctgc
26951 gagcaggccc cggaaaccga agccccgcc gcccggccg ccgagccggc
27001 gccggaggcc acgcccaccg tcgtccctg ggtcgtctc ggacgcagcc
27051 gggaagcgt ccaggcgag ctggaccggc tcaccgcga caccgccgc
27101 caccgccgc gtcggcggc ggacgtcggc cgtcgctgg ccaccgacc
25 27151 cacgtcttc ccgcaccgc ccgtgctgt cggcgcccg gacggggtgc
27201 gcgaggcgc ccgcgcgcc gcgcccga ccccgggc caccgcgttc
27251 ctgttctccg gacaggcgc ccagcacgc ctgatgggc acgacctgta
27301 ccagcgcttc ccgtctacg ccgacgcact ggacaccgt ctcgccagt
27351 tcgacaccgt gctggacgt ccgctgcgc ccgcgtgtt cgccgcgcc
30 27401 ggcaccccg aggcgcgct cctggaccag accggcttca cccagccgc
27451 gctgttcgcc gtcgaggtc cactgttccg gtcgccgag tcctggcggc
27501 tgacgccgga cttcgtcgcc ggccactcca tcggcgagat cgccgccgc
27551 cacgtcgcg ggggtgttc cctggaggac gcctgcacg tggtcgcgc
27601 ccgcgcctcc ctcatgcag aactgccgc ggacggtgc atggtggccc
35 27651 tggaagccac cgaggacgag gtcgcgcgc tgcaccca cggcgtcgca
27701 ctcgcgcgg tcaacggccc ccgctcggg gtcgtcgcg gcgccgagga
27751 cgccgtccgc gcggtcgcg accggctcgc cgccgacgg cgccgacccc

27801 gccggctgac ggtcagccac gccttccact cgccgctgat ggacccgatg
27851 ctcaccgact tcgcccgggt cgcgaggggc ctgacctacc acgagccgcg
27901 catccccctc gtctccaccc tcctcggcgc cccggccggc gccgaactgc
27951 gcacccccga ctactgggtg cggcacgtcc gcgagaccgt gcggttcgcc
5 28001 gacggcgtgc gcgccttga cgacgccggc gccggcacct tcgtggagat
28051 cggcccggac ggctgtctca ccgccctgac ccagcagacc ctcgacaccg
28101 tcgaggccgg cgcgcccgcc gtcgtcgtgc cgctccagcg ccgcgaccgc
28151 gccggcgacc tcgcctcctt ggagggcctg gccaccctgc acaccacgg
28201 caccggcccg tcctggcccg cctacttcca ggccaccggc ggccaccgga
10 28251 ccgatctgcc cacctacgcc ttccagcggg agcggtaactg gccgaactc
28301 ggcgaccccg tcgccaccgc cccgcaggac ccggcgccct ggcgctacca
28351 cgagacctgg gccccgctgc cggcacccga ggcgcccgcg ccggccggcc
28401 gcgccttggc cctcgtcccg gccgggaacc gcgacaccgc gtggatgacg
28451 gccgtcgccg acgcgtcggc cgccgacacc gtcacggccg aaccggacgc
15 28501 cctggccgag cagttgacgg ccgccggcga cacaccctgg cgcgctcgtg
28551 tgtcgtgctt cgccgcccg tccgaggggc tgcccgcgga cggcgccctg
28601 cccgcgcgcc tcctcgccac cctggacgag gccggcgtgc acgcgcccct
28651 gtggtgcgtc acccgcgcg ccgtcgcggt cgcgggggag gccccgaccg
28701 ccgtcggcca ggccgccctg tggggcctgg gccgggtcgc cgcgctggac
20 28751 cacccggaac gcttcggcgg cctggccgac ctgcccggcg acaccgacgc
28801 gcacgccgcc gggtgtctcg ccgcgcacct ggccgcgccg ggcaccgagg
28851 ccgagatcgc ggtccgcgcc accggcgtcc acgcccgtcg cctggtccgt
28901 acgccggccg ccgccgacgg tgccacctgg ctgccgaccg gcaccgtcct
28951 ggttgtcggc ggcaccggcg gcaccggcac catgggcggc cgggcccggc
25 29001 gctggctggt ccgcgagggc gcccgccacc tcgtcctgac cgccccgac
29051 ggcaccacga ccgcccgga caccgaggcc ctgacggccg aactggccgc
29101 gctcggcgcc cggatcacgg tcgtggacca cgaccccacc gccccggacg
29151 gcttcgccgc gctcctcgac ggactgcccg acgacacccc gctcaccgcg
29201 gtcgtgtacg cgccggaggc cgacgccgcc cccggcaccg cggccgagct
30 29251 gtccgcgca ctgcccccg tcaccgccct aggcgccgcc ctcaccggcc
29301 ggccgctgga cgccttcgtc ctcttcgggt ccategccgg gctctggggc
29351 gtgcgcggcc gggccgccga ggccgcgtcc ggcgccctacc tcgacgcctt
29401 cgcccgccgc tgccgcgacc gcggcacccc ggcaactggc gtcgcctggg
29451 gcgcctgggc cgacctggtc ggcccgtccc tcgcccgcga cctgcggatg
35 29501 aacggcctgc cgggtgatga cgccgacacc gcaactgacc ccctcagccg
29551 ggccgctgcc gacgggtccg ccgccgaggc ggtcgccgac gtccgctggg
29601 agaccttcgc gccctccac cacgaggccc gccgcaccgc cctgttcgac

29651 gccctgcccg aggccgcgcg cgcgctcgcg gaggccgccc gggaccgcgc
 29701 cgaccggaag accgccgccc gcgactacgg ccggtggctc gccgagcagc
 29751 ccgccgcgga ccacgacgcc atcctgctgg cactggtcac cgagaaggcc
 29801 gcgaccgtcc tcggccacgc cgaccacgac ctgctcgaac ccgacctgcc
 5 29851 cttccgcgac ctgggcttcg actcgtgac cgcggtcgac ctgcgcaacc
 29901 agctcaccgc ggaaaccggc ctcaccctgc ccgccaccct cgtcttcgac
 29951 caccccaacc cgcccgccct cgccgcccac ctgcgcgccc aactcctcgg
 30001 cgaggcgagc gactccgccc caccggtggc cgcgcccgtc gccctcgggtg
 30051 ccgacgacga cgcgatcgtc atcgctcgga tggcctgccg ctaccccggc
 10 30101 ggggtcacct cgcccagga cctgtggcag ctggtcggcg acgaggtcga
 30151 cgcggtcggc gacttcccga ccgaccgcgg ctgggacctg gccgcgctcg
 30201 ccggcgacgg accggggcgc agtgccaccg ccagggcggt attcctctac
 30251 gacgccaccg acttcgaccc cggcctgttc ggcatctcgc cgcgcgaggg
 30301 cctggtgatg gaccgcgagc agcggatcct gctcgaaacg tcctgggagg
 15 30351 ccctggagcg ggccggcatc gaccggcgga cgtgcgcgg cagcggcacc
 30401 accggcgtct tcgtcggcgg cggctccggc gactaccggc cgccggagga
 30451 ggccgggcag tggcagaccg ccagtcgc cagcctgctc tccggtcgcc
 30501 tcgcctacac cttcggcatc cagggcccca ccgtgtcggg cgacaccgcc
 30551 tgctcctcgt cgctggtcgc gctgcacctg gccgcgcagg ccctgcgcgc
 20 30601 cggcgaatgc tcgatcgcg cggccggcgg cgtcaccgtg atggccaccc
 30651 cgggtggcctt cgtcgagttc agcgcaccag gcgccctgtc gccggacggc
 30701 cgctgccgcg cttctccga cgacgccaac ggcaccggct ggtccgaagg
 30751 cgtgggcatg ctctcgtcgc aacggctctc cgacgccgc cgcaaccggc
 30801 accgcgtcct cgccgtgctc cgcggctccg ccatcaacca ggacggcgcg
 25 30851 tccaacggcc tgaccgcccc cagcggcccc gcccagcagc gcgtcatccg
 30901 ccaggccctc gccaacgccc gactgcgccc cgccgacatc gacgccgtcg
 30951 agggccacgg caccggcacc aggtcggcg accccatcga ggcccaggcc
 31001 ctgctcgcca cctacggcca ggaccgcgag cggcccgtgc tgctcggctc
 31051 gctcaagtcc aacatcggcc acaccaggc cgcctccggc gtcggcgggc
 30 31101 tcatcaagat ggtcctcgcc atgcagcacg gcgaactgcc gcgtccctg
 31151 tacgccgaga acccctcgtc gcacgtggac tggaccgccg gccgcgccc
 31201 cctgctcacc gccaggaccc cgtggcccga ctccggtcgg ccgcgcgcg
 31251 ccgccgtctc ctcttcggc gccagcgga ccaacgccc cgccatcctg
 31301 gagcagccgc cgcgcgagga actccccgc cgcgcccgcg acgacggcgc
 35 31351 cccgctgccg ttcttgctct ccggccgctc gcagaacgcc ctgcgcgccc
 31401 agggccgccc actcctggcc cgcctaccg cccacccgga caccgggccc
 31451 gccgacctgg cgtactccct ggcgaccacc cgggcccgcct tcgagcaccg

31501 ggccgcgatc accgccaccg accacgacgg cctccgcacc ggcctgaccg
31551 ccgtcgccga gggcaccacc gccccgcaca ccgccgaaca ccacctccag
31601 ggcaccgga agcgcgccgt gctcttctcc ggccagggct cccagcgctt
31651 gggcatgggc cggaactgc acgagcgcca cccggtgttc gccgagcggt
5 31701 tcgactccgt actggccgc ctcgacgacc ggctcgacac cccctgctgg
31751 gacgtcgtct ggggaccga cgaggaggcg ctgcacgcca ccgggaacac
31801 ccagcccgcc ctgttcgccc tcgaagtcgc gctctaccgc ctgatcgaat
31851 cctggggcgt ggggcccgc ttcgtggcgg gccactccgt cggcgagctc
31901 gccgcggccc acgtcgccgg ggtgctctcc ctggacgacg cctgccgcct
10 31951 ggtcgcccgc cgcgccgccc tcatgcagcg cctcccggcc ggcgcgcca
32001 tgatcgccgt cgaggccacc gaggacgagg tcaccccgct cctcaccgac
32051 ggcgtgtccc tcgcccggt caacggaccg accgccgtgg tctctccgg
32101 cgcgggcgac gccgtgaccg ccctgggcca ggcgctggcc gaacggggcc
32151 accgcaccac ccggtcgcgg gtcagccacg ccttccactc gcacctcatg
15 32201 gaccgatgc tggcggaactt ccgcaccgtc gccgagggcc tggaatacca
32251 cccgcgcgc atccccgtgg tctccaacct caccggggac gtcgcccagc
32301 cgcccgacct gtgctccgcc gactactggg tgcgccacgt ccgcggcacc
32351 gtacggttcg ccgacggcgt gcgcaccatg gccgaccgcg gcgtgcacct
32401 cttcctcgaa ctcggcccg acgccgtgct gtcggccatg gcccgccagt
20 32451 gcgcaccgga cgccgtcgtc gtcccggccc tgcgccgcaa ccgcgacgag
32501 gacgagacgc tggtcggcgc cgtcgcgca ctgcacgtcc acggcgcggg
32551 tccgcgctgg gacgcgtact tcgccggccg cgcgcccag tggctggacc
32601 ttccgacgta ccccttccag cgcgcccgct tctggccgga gtcccttccg
32651 ggcgccgcat cggccgcccc ggcagccgga cagccggccg agaccgacgc
25 32701 ggccttctgg gacgccgtcg cacaggagga cttcaccgca ttggaatccg
32751 tactcgacgt cgagagcgac gcaactgtcca aggtgctgcc ggccctgatg
32801 gactggcgca gccgccaggc cgacgagtc caactggcag gctggcgcca
32851 ccgcatcgtc tggaagcggc tcaccggcgc cgccctggca caccgcaagg
32901 cgctcagcgg cacctggctc gcggtggtcc ccgagggctt cgccgacgac
30 32951 ccctgggtga ccaccacct ggacggcctc ggtaccacc tcgtgcatct
33001 ggaggtcgcg gaggccgacc gggccgcgct ggccgacgcg atcgcgccc
33051 gcaccgcca cggcaccgc ttcggcgcg taatctccct gctggccctg
33101 cgcgaggagc tcaccggcgc ggtgcccag gggaccgcc tgaccaccac
33151 cctcctccag gccctcggcg acgccggcgt cgacgcaccg ctgtggtgcg
35 33201 tcaccgcag cgccgtctcc gccggccgca ccgaccggcc gcaccgaccg
33251 ctcaaggcg ccgtctgggg cctgggccc gtcgcgccc ttgagtacc
33301 gcagcgctgg ggcggcctgg tggacctgcc ggaggagccc gacgagcggt

	33351	ccgcggcccg	cctcgccgcc	gtcctggccg	gtctggacgg	cgaggaccag
	33401	gtcgccgtgc	gcggcaccgc	ggtgctcgcc	cgccgcctgg	tgccggctcc
	33451	cgcccgcaag	ccgtcccggc	cctggcacc	gtccggcacc	gtcctggtca
	33501	ccggcggcac	cggcgccctc	ggcgcgcacg	tcgcccggcg	cctggccaag
5	33551	gacggcgccc	agcacctcgt	cctgctcagc	cgccgcggcc	cgagcgctcc
	33601	cggtgcgggc	gaactgcgcg	cggaactgga	cgcggtgggc	accgacgtca
	33651	cggtcgccgc	ctgcgacgtc	gccgaccgcg	accagctgac	ggccgtcctg
	33701	gacgcgctgc	ccgccgaccg	gccgctgacc	ggtgtggtgc	acaccgccgg
	33751	cgctcctcgac	gacggcgtag	tggaccggct	cacccccgag	cggttcagg
10	33801	aggtgttccg	cgccaaggtc	acctcggcc	tgctgctgga	cgagctgacc
	33851	cgcgaccgcg	agctggccgc	gttcgtcctc	ttctcctccg	cctccgccgc
	33901	ggtcggaac	ccgggccagg	ccaactacgc	cgctgccaac	gccgtcctgg
	33951	acgcgctcgc	cgaacagcgc	cggtgtctcg	gcctgccgcg	cacctcggtc
	34001	tcctgggggtg	cctggggagg	cggcggcatg	gccgacgccg	acggcgcgga
15	34051	cgaggccgcc	cggcgcggcg	gcgtcggcgc	catggaccgc	cacctcgccg
	34101	tggaagccct	gctgcgcctg	gtcgccgaga	aggagccgac	cgcggtggtc
	34151	gccgaggtgg	ccctggaccg	gttcgccggc	gccttcggcg	gcagccgacc
	34201	cagcgccctg	ctgcgggagt	tccccgcta	ccgcgaggcg	ctcgccgcc
	34251	aggcggagca	ggccgcggac	ggcggcgggc	tgcccgcccg	actggccgcg
20	34301	ctgcgcgccg	cccgcgcct	ggacaccgtt	gtggacctgg	tgcgccaccg
	34351	cgccgcgcag	gtgctcggt	accccgacac	cgaagcggtc	gccgccgaac
	34401	ggtccttccg	cgacctgggt	gtcgactcgc	tcggcgccgt	cgagctgcgc
	34451	aaccaactga	gcgcggccac	cggcctgaac	ctgccggcga	cgctgggtgt
	34501	cgaccaccgc	acccctctgg	tcctggggga	gcacatcctc	ggcgggctct
25	34551	tcccgagcga	gcccgcgggg	tccgacgacg	agacggagat	ccggggcctg
	34601	ctggcctccg	tccgcgtcga	ccaactgcgg	gagatcgggg	tcctggagcc
	34651	cctgctccag	ctcgccggac	gcggcgggcg	ggccgcggac	ggcgacgacg
	34701	gcgagtcctg	cgactcgatg	acagtggcag	acctggtgcg	ggccgcgctc
	34751	aacggccagt	ccgacctgta	gcgcgattga	tgagacgac	gatgaacgcg
30	34801	cccgagaacc	ccgagacccc	cgagaacaac	gtagtcgccg	cactccgcgc
	34851	cgcggtcaag	gagaccgacc	ggctccggcg	gcagaaccgg	atgctggtcg
	34901	cgcgggccaa	ggaaccgatc	gccgtggtcg	gcatggcctg	ccgcttcccc
	34951	ggcgccgtcg	actccccgga	agcgctgtgg	gagatggtcg	ccaccggcac
	35001	cgacgtgatc	tccggattcc	ccgacgaccg	cggtggggac	ctggaggcgc
35	35051	tgcgcaacag	cggcaccgac	gcccgcgaca	ccgacgtcag	ccagcgcggc
	35101	ggattcctgg	actgcatcgc	cgacttcgac	cccggcttct	tcgggatctc
	35151	accgcgcgag	gcggtcacca	tggaccgcga	acagcggctc	ctgctgacca

35201 ccgcctggga ggccgtcgag cgggccggca tcgacgccac cacgctgcgc
35251 gccacccgca ccggcgcggt catcggcacc aacggccagg actacgccta
35301 cctgctcgtc cgctccctgg acgacgccac cggcgacgtc ggcacccgca
35351 tcgccgccag cgccgectcc ggtcggctct cctacacctt cggcctcgaa
5 35401 ggccccgcgc tcaccgtcga caccgcctgc tcctcgtcgc tggtcgcctt
35451 gcacctggcc gtgcaggcgc tgcgcaacgg cgagtgcggc atggcgctgg
35501 ccggcggcgt caacgtgatg gccacaccgg gctcgttgtt cgagttcagc
35551 cgccagggcg ggctggcccg ggacggccgc tgcaaggcgt tcgcggacgc
35601 cgccgacggc accggcttgt ccgagggcgc cggcggtgtg ctgctggaac
10 35651 ggctctccga cgcacagcgc aacggccacc cgggtcgtgc cgtggtccgc
35701 ggctccgcgg tcaaccagga cggcgctctc aacggcttca ccgccccaa
35751 cggccccctc cagcagcgcg tcatccgcca ggccctcgcc aacgcgggcc
35801 tggccaccgg cgacatcgac gcggtcgagg cgcacgggac cggcaccctg
35851 ctcggcgacc ccatcgaggc gcagagcatc ctcgccacct acggccagga
15 35901 ccgcgcccac ccggtgctgc tcggctcgat caagtcgaac atgggccaca
35951 cccaggccgc gtccggcgtc gccggcgtga tcaagatgat catggcgatg
36001 cggcacggcg tcctcccgcg gaccctgcac gtcgaccggc cctccaccca
36051 cgtcgactgg accaccggca gcgtcgaact cctcaccgac gccacaccgt
36101 ggccccgagc tggcaggccg cgccgcaccg gcatctcctc cttcggcgtc
20 36151 agcggcacca acgcccacgt catcgtcgaa caggcccccg acacccccgc
36201 cgaggcggct gacgacactc cgccccgcac cccgcggacc ctgccgtggc
36251 tgctctccgc ccgacccggc gccgccctgc gcgaccaggc caccgcgctg
36301 ctcgaccacc tcgaccgcc cgcggcgac cgcgggccca ccgcctgga
36351 caccgcgttc tcctcgcga ccaccgcgc cgccctggaa caccggctcg
25 36401 ccgtcgtcac cggcaccgac ggcaccgccg gacgggacgc cctgaccgcc
36451 tggctggcgc acggcaccgc ccccgacgcc cacgaaggac acgcccggg
36501 acgcaccgcg tgcgcggccc tcttctccgg ccaggcgccc cagcgctgg
36551 gcatgggccc cgaactccac gcccggttcc cgggtgtcgc acgggccctc
36601 gacaccgccg tggacctgct cgacgccgaa ctgggcgga ccctgcggga
30 36651 ggtgatctgg ggcaccgacg acgcgcgct caacgagacc ggcttcaccc
36701 agccccccct gttcgcgctc gaggtcgccc tctaccgctt gatcgaatcc
36751 tggggcgctc ccccgactt cgtcgccggc cactccatcg gcgagatcg
36801 cgccgcgcac gtcgccgggg tgttctccct ggaggacgcc tgcacgctgg
36851 tggccgcccg cgccgggctg atgcaggcgc tgccgcggc cggggcgatg
35 36901 gtcgcgctc agggcaccga ggacgaggtc agcccgctgc tcaccgacgg
36951 cgtcgcgatc gcccgatca acggccccac ctcgctcgtc gtctccggcg
37001 acgagaccgc caccctcgcc gtcgccgcc gactcgccga acagggccgc

37051 cgcaccaccc ggctgcgggt cagccacgcc ttccactcgc cgctgatgga
37101 cccgatgctc gcggagttcc gcgcggtcgc cgagggcctg tcctacggcg
37151 aaccgcagat cccggtggtc tccaacctca ccggcgcggt cgccgacggc
37201 accctgctcg gcactgccga ctactgggtc cggcacgtcc gcgaggcggt
5 37251 ccgcttcgcc gacggcatcc gcgccctcac cgacgccggc gtcggcgcct
37301 tcctcgaact cggcccggac ggcacgctcg ccgccctggc ccagcagtcc
37351 gccccgcagc ccgtctccgt ccccgctcctg cgcaaggacc gggacgagga
37401 gcccgcgcgc gtcgccgcac tggcccggct gcacaccgcc ggcgtcccg
37451 tggactggac ggcgttctac gccggcaccg gcgcccaccg caccgacctg
10 37501 ccgacctacg ccttccagta cgagcgctac tggcccaagg ccacctaccg
37551 gcccgcgcgc gccaccggcc tcggcctgac cgccgcgcgc caccgcgtgc
37601 tcggcgccgc catgtccgtc gccgggtccg acgagctcct gctcaccggc
37651 accctgtcgc tcgccacca cccctggctc gccgaccacg tcgtcggcgc
37701 catggtcttc ttccccggca ccggcttccg ggaactggcg gtccgcgcgc
15 37751 ccgaccaggt cggctgcgac cgggtcgagg aactcatgct cgccgcgcgc
37801 ctgatcctgc ccgccaccg caccgtccag atgcagatcg cggtcggcgc
37851 cgcggacgac gacggcggcc gcgacctgcg cttcttcacc cggcccgggg
37901 acgaccggga cgccgcctgg gcccagcacg ccaccggccg gatcaccgag
37951 ggcgagcgcg tcctcgccct cgacaccacc acctggccgc cccgcgacgc
20 38001 ggaaccgcgc gacatcgacg gcctctacga ccgctaccgc gccaacggac
38051 tcgactacgg gcccgctctc cgcggcctgc gcgccgtatg gcgccgcgac
38101 accgagatct acgccgaggt cggcctgcc gaaggcaccg ccgacgccga
38151 cgccttcggc ctgcaccgg cctcttcga cgccgtcctg cacagcacc
38201 tcttcgcctc cgccgacggc gacgaccgca gcctcctgcc gttcgctgg
25 38251 aacggcgtgt ccctgcacgc cgcgggcgcg gacgcgctgc gcgtccgat
38301 caccagctgc ggccccgacg ccgtggagat caccgcgcgc gaccgcagg
38351 gccgccccgt cgtctccgtc gaatcgctga cgctgcgcgc cgccggcccc
38401 gacgccggca ccgccacca ccgtgccgac gcgggctccc tcttcgcgat
38451 ggactggacg cccgcacccg tccacgcccc ggccaccccc gccacctggg
30 38501 ccgtcctcgg taccgacctg atcggcctga ccgaggcgct caccgcgcgc
38551 ggccccgaca ccgtcacggg actccgcgac ggtgtcgacg ccctcggcga
38601 actcaccgcc ggcgacgacc gcccggtgcc cgacgtggtc gcggtaccgc
38651 tgcgcggcgc caccgaccac gggccggccg gtgcccacga cctgaccgc
38701 accgtcctgg ccctgctcca ggaatggctg gccgaggagc gcttcgccc
35 38751 ctccccggtg ctgctcgtca cccgcggcgc ggtcgccgac ggcgagcgcg
38801 gcccgcctga cctggccgcc gccccggtct ggggcctggt gcgtccgcc
38851 cagtccgaga accccggccg actgctgctc gtcgacctcg acgacaccgc

	38901	cgagtcgcc	gcccactcc	cgttgctgcc	ggccctcctg	gacgccgacg
	38951	agccgcaggc	cgtgggtccg	gagggcaccg	tccgggtcgg	ccggctcgcc
	39001	cgctggact	ccggccgagg	cctcgtccc	ccgcccggca	ccccctggcg
	39051	cctgggcagc	cgggccaagg	gcagcctcga	cgccctcgcc	ctgctgcccc
5	39101	accccgaggc	ccgacgaccc	ctcaccggcc	acgaggtccg	cgtcggcatc
	39151	cgcgccgagg	gcctgaactt	ccgtgacgtg	ctcaacgcgt	tggggatgta
	39201	tcccggggat	gcggggctgt	tcgggttcgga	ggcggccggg	gtggctcgtcg
	39251	aggtcggacc	ggaggtcacc	ggcctggcac	ccggcgaccg	ggtcattgggc
	39301	atgctcttcg	gcggcttcgg	accgctcggc	atcgccgacg	cccggctgct
10	39351	cacccccgtc	ccggccgact	ggctcctggga	gacgggtgcg	tcgggtgccgt
	39401	tgggtgttct	caccgcgtac	tacgccctga	aggagttggg	tgggtcgcgg
	39451	gcgggggaga	aggtgctggt	gcattgccgt	gccggtgggt	tcgggtatggc
	39501	ggcgatccag	atcgcccgcc	atgtcgggtg	cgaggtgttc	gccacggcca
	39551	gtgagggcaa	gtgggacgtg	ctgcgctccc	tgggcgtggc	cgacgaccac
15	39601	atcgctcct	cccgaccct	cgacttcgag	gcggccttcg	ccgaagtgcg
	39651	cgcgaccgc	ggcctggacg	tcgtactgaa	cgcgctgtcc	ggcgagttcg
	39701	tcgacgcctc	gatgcggctg	ctcgcgacg	gcggccgggt	cctggagatg
	39751	ggcaagaccg	acatccgcgc	cgcggaactc	gttcccgacg	gcctctccta
	39801	ccactccttc	gacctcggca	tggtcgatcc	ggaacacatc	cagcggatgc
20	39851	tgctcgacct	cgctcgagctg	ttcgaccgcg	gcgcgctggc	cgcggtgccg
	39901	gtccgcagct	gggacgtgag	ccgcgcccgc	gagggcttcc	gcttcattgag
	39951	cctggcccag	cacatcggca	agatcgtgct	caccgtgccg	caaccctcgc
	40001	accccgaccg	caccgtgctc	ctcaccggcg	gcaccggcgg	cctggccggc
	40051	ctgctcgccc	gccacctggg	caccgagcac	ggcgcccggc	acctgctgct
25	40101	ggccggcccg	cgcgggcccc	acgcgcccgg	cgccgcccga	ctccacgccc
	40151	aactgaccgc	cctggggcgc	gaggtcaccg	tcgccgcctg	cgacgtcgcc
	40201	gaccgcaccg	cgctcgccgc	gctgctcgcc	accgtgcccg	ccgaacaccc
	40251	cctcaccgcg	gtcgtgcaca	ccgcccgggt	cctggacgac	ggcaccctca
	40301	ccgccctgaa	ccccgaccgc	ctcgccaccg	tcctacggcc	caagggtggac
30	40351	gccgcctggc	acctgcacga	cctcaccgcg	cacctcgacc	tggccgcgtt
	40401	cgtgctctac	tcctccaccg	ccggcgctcat	ggcgggaccg	ggccaggcca
	40451	actacgcggc	cggaacaccc	ttcctcgacg	cgctcgccgc	ccaccgacac
	40501	gccctcggcc	tgcccggcac	ctcgctggcc	tggggcgcc	gggagcaggg
	40551	cgccggcatg	accggcgcac	tgaccgacca	cgacctgcgc	cgggctcagcg
35	40601	acgcggcgcg	ccaaccgctg	ctcaccggcg	aacgcggcct	cgccctctac
	40651	gacgccgcca	ccgccgcca	cgaaccctg	atcgctccgc	tcggcctcac
	40701	cgcggtgag	ctgcccggcg	gggtcggcgt	ccccgcccgt	ctgcgcccgc

40751 tgggtccgcac cgcggggccgc cgggccaggc cgggcaccgc cggcgtctcc
40801 cgcgcgggcc tcgccgaacg cctcgccgcc ctgcccaggc aggagcgcac
40851 ccccttcctc gtcgagctgg tgcgcaccga ggccgccacc gtcctcggcc
40901 acggctccac cgaccgggtg gacgcccgcc gcgagttccg ccaactcggc
5 40951 ttcgactcgc tgaccgccat cgaactgcgc aaccgactcg gcaaggccac
41001 cggcctcacc ctgcccgcc cctcatctt cgactaccgc acccccgacc
41051 gcctcggcgt ccacctccac gacgaactcc tcggcgcgga cggcccggtg
41101 accgtcaccg ccgccgcaca ggccgcggac ccggagcacg acccggtcgt
41151 catcgtcggc atgagctgcc gcttccccgg cggcgtcagc tccccgagg
10 41201 agctgtggga cctggtggca tccggcaccg acgcgatcac cggcttcccc
41251 gccgaccgcg catgggaccg ccaccgcag ctgcgcggcg ccccggcgcg
41301 ccgcaccggc caggggcgat tcttcgcga catgcgcgac ttcgacgcg
41351 ccttcttcgg catctcgccg cgcgaggccc tggccatgga cccgcagcag
41401 cgcatectcc tcgaagtgc ctgggaggcc gccgagcgcg ccggcatcga
15 41451 cccgcagacc ctgcgcggca gcgacaccg cgtgttcatt ggctcagcg
41501 gccaggacta cggcgccctc gtgatgcgt cccgcgacga catgcgcggc
41551 cagccacca cgggcctcgc cgtcagcgtc gtctccggcc gcctcgccta
41601 cgcgctcggc ctggagggcc cggccctgtc cgtggacacc gcctgctcct
41651 cctccctggt gtcgctgcac ctggccgccc aggcgctgcg cgcgggggag
20 41701 tgcaccatgg ccctggccgg cggcgtcacc gtcattacca ccgccgcaa
41751 cttcaccggc ttctcccgga tgggcggcct cggccaggac ggccgctgca
41801 aggcgttctc cgactccgcc gacggcaccg gctggtccga gggcgccgcc
41851 gtctggtcc tggaacgcct ctccgacgcc cggcgcgccg gccaccgcgt
41901 actggccgtg gtgcgcggct cggcggtcaa ccaggacggt gcgtccaacg
25 41951 gtctgacggc gcccaacggt cccgcccagc agcgcgtcat ccggcaggcc
42001 ctggccaacg cgggcctgac cccgctcgac gtggacgccg tcgaggcgca
42051 cggcaccggc acccgctcg gcgaccccat cgaggccag gccctgatcg
42101 ccgcctacgg caccgaccgc gacccgaac acccgctgct gctcggctcg
42151 gtgaagtcca acatcgcca caccagtcg gcggccggcg cggccgggct
30 42201 ggtcaagatg gtcattggca tgcgccacgg catcctgccg cagacctgc
42251 acctcaccga accgtcctcg cacgtggact ggtcggcggg cacggtgcgg
42301 ctgctcaccg agcggaccgc ctggcccgcg acggatcgtc cgcgtcgggc
42351 cggggtctcc tcgttcggca tcagcggcac caacgcccac gtcactctgg
42401 aacagccgcc cgcgagccc acccccgccg ccgacccggg ccggcccgcga
35 42451 cccaccgtgg tggcctggcc cgtctccgcg cagaccccg ccgccctcga
42501 cgcccaactg gaccggttgc gcaccgccgc cgcctggcg ccgctcgaca
42551 ccgcccacac cctcgccacc ggccgctcgc tcttcgaaca ccgcgcgtc

42601 ctgctcgcca ccgtcggcga cccggcgacc ggcgccccg acctgcccga
 42651 ggtcgccagg ggagcggcga cgcgcacccg caccgcgttc ctcttctcgg
 42701 ggcaggggtgc tcagcggtcg gggatggggc gtgaactgca tgctgcttcc
 42751 ccggtgttcg cggcggcggtt cgacgaggtg gtggctgtgt tggatgcgga
 5 42801 gctgggttct gatgctgatg ggggtgtgtc gctgcgggag gtgatgtggg
 42851 gcgggggggtc ggagtgtgtg gatcgaacgc gtttcacgca gccggcggtg
 42901 ttcgcgggtgg aggtggcggtt gttccgtttg gtggcctcgt ggggggtggg
 42951 gcctgagttc gtggcggggc attcgggtgg tgagattgcg gcggcgcatg
 43001 tggccgggggt gttctcgttg gtggatgcgt gtcgtttggt ggtggcgcg
 10 43051 gcttcgttga tggatgcgtt gccggtgggt ggcgtgatgg ttgcggtgga
 43101 ggcggccgag gcggaggtgg tgccgctgtt ggtcgatggg gtggcgatcg
 43151 ccgcggtcaa cgggccggtc tcggtgtgtg tctccggtgt ggaggcggcc
 43201 gttgggcagg tcgtggatca gttgtgtggag cggggccggc gggccgctcg
 43251 gttggcggtc agtcagtctt tccactcgcc gttgatggat ccgatgttgg
 15 43301 atgccttcg gccgctcgcc gagggcctgg agtaccacca gccgcgcatc
 43351 cccgtggtgt ccaacgtgac gggcgaggtg gccgcggcgg aggagctgtg
 43401 cgcggccgac tactgggtgc ggcacgtccg ggcgaccgtg cgttcgccg
 43451 acggcgctcg caccctggcc gagcgcgcg ccaccgcctt cctggagatc
 43501 ggccccgacg gcgtactgtc cgcgctcgcc cgcggcgctc tgcccccgga
 20 43551 ggcgctcgtg acgcccaccc tccgcaagga ccgcgacgag gagagcgccc
 43601 tgctcgccgg actggcccgg ctgcacgtcg ccggcgtagc cgtcgactgg
 43651 agcgccgccc tgaccggcac cggcgcgcg gcgaccgacc tgccgacctg
 43701 cgccttccaa cgcgagcgg actggccgga gttggccgcc gaaccgcggg
 43751 gcggcgcgcg ggatgccgag gacgcggagt tctgggcccgc ggtggagcgc
 25 43801 gcggacgcca ccgcgtcgc cgcacacctg gacatcgacg gcgaccagct
 43851 cggcgccgtg ctgccgcac tgtccgcctg gcgcaccgg cgccgcacca
 43901 catcgcccac caacgccctg cggcaccggg agagtggga accgctgtcg
 43951 ctgccgggca cgcgcacac cggcgcggtc ctggtgctgg tgcccccgcc
 44001 cgcgaccacc gaccctggg tcgccgacgt cgtcgccgcc ctgcggccgg
 30 44051 acgcccgcg ggtggacgtc ccggccgacg gcaccgaccg ggccgcgtc
 44101 gccgccctgc tcaccgaagc ggccgacgac accgccccga ccgccgtggt
 44151 ctccctgctc gcgctcgacg agaccagcgg cgacgacgag gtaccggccg
 44201 gcaccaccgc caccgcccgc ctctgccagg ccctcgccga caccggcgcc
 44251 ccggccccgc tgtgggccct gaccggggc gcggtcgccg cgctccccga
 35 44301 cgagcagccg accgccccg cccaggccgc cgtctggggc ctgcggccga
 44351 tcgcccctc cgaactccg cgcactggg gcggactggt cgacctgcc
 44401 gccgacctc acgagcgac cgcacgccga ctgcccgcg cactggccga

44451 cgccggcgac gaggaccagc tcgcgctgcg ggccaccggc gcctacggcc
44501 gccggatcac cccggcgccc gcccccgacg acgcccccg caccgggtgg
44551 cagccgaccg gcaccgtcct gatcaccggc ggcaccggcg cgctcggccg
44601 gcacaccgcc cgctggctcg ccgccacgg cgccgagcac ctgctgctgc
5 44651 tcagccgcag cgccccgac gcgcccggcg cggccgaact caccaccgaa
44701 ctaccgccc tcggcgccc cgtaacctc gtggcctgcg acgcgcgga
44751 ccgggagcag ctgaccaggg tcctcgccga ggtaccggcg gactgcccgc
44801 tgaccggcgt ggtgcacacc gccggagtgc tcgacgacgg cgtgctcacc
44851 ggctcaccc cggaccggtt cgccacggtc ttccgcgcca aggtggcctc
10 44901 cgccgtgctc ctgacgagc tgaccggga caccgacctg gccgtcttcg
44951 cgctgttctc gtccgtcgcg ggcgctgctg gcaaccccg ccaggccggc
45001 tacgcccgcg ccaacgcggt cctcgacgcc ctcgccgccc gccgcgggc
45051 ccagggcctg gccggcacct cgatcgctg gggcgcttg gccggcgacg
45101 gcatggcggc ccgtcacacc cgcgccggcg ccgaaccctg cggcctgctc
15 45151 gaccccgacc tcgccgtacc ggccctggcc cgcgcggtga cggagcccca
45201 gccaccctc gtctcgccg acctccagca gcccgggctg ctggaatccc
45251 tgctggcgct gcgcccagc ccgtctctga gccggtgccc cgcgcgccgc
45301 accgcggccc gcgcggtcca ggaagcggac cgcgcgag ccggcgccgc
45351 ggccgacctg cgcgaccaac tcgccggcac cgcacccgcc gaccgccacg
20 45401 ccgtcctcct ccgcctggtg cggaccacgg ccgcgcggt cctcggccac
45451 accggtgccg acgccatccg ggccgacaag cccttcgcg acctcgctt
45501 cgactcgctc accgcggtgg aactgagcag cgctctcgcc gccgccaccg
45551 gcctggccct ccgcccagc ctgctcttcg accaccctc ccgcggggcg
45601 ctgcgcgacc acctgcgcg cgaactcacc ggcgaccggc cggaatccgc
25 45651 cccgcggccc ccgccggcac cgtccccgc cgcggacgac gatccgatcg
45701 tcgtcgtcg catggcctgc cgcttccccg gcggcgctac caccgccgag
45751 gagttctggc agctgctgc cgaggccgg gacggcatcg acgcgttccc
45801 caccgaccgc ggctgggacc tcgacgtgct cggccggcga cggccagggc
45851 cacagcggcc accgaggtcg gcggcttcct cttacgacgc ggccgccttc
30 45901 gaccccggtt tcttcgacat ctccccgcg gaggcgctcg ccatggaccc
45951 gcagcagcgg ctgctgctgg agaccgcctg ggaggccgtc gaacgcaccg
46001 gcaccgaccc gaccggctg cgcggcagcc gcaccggcgt gttcgtcggc
46051 accaacggcc aggactacgc cggcctcgtc ctgcgcgccc aggaggacgt
46101 cgaggggcac gccggcaccg gactggccgc cagcgtgatc tccggccgcc
35 46151 tcgcctacgc cttcggttc gagggcccc cggtcaccgt cgacaccgcc
46201 tgctcctcct cgctcgctgc cctgcactgg gccgtccagg cgctgcgcgc
46251 gggggagtgc tccctggccc tggccggcg cgtaaccgtc atgacgaact

46301 cgacgagctt cgccggcttc acccggcagg gcggcctggc gccggacggg
46351 cactgcaagg cgttctccga ctccgccgac ggcaccggct ggtccgaggg
46401 tgtgggcgtc ctcgctcgcg aacgccgctc cgacgcgctc cgcaacggcc
46451 atgagatcct ggccgtggtg cgccgctcgg cggtaacca ggacggtgcg
5 46501 tccaacggtc tgacggcgcc caacggtccc gccagcagc gcgtcatccg
46551 gcaggccctg gccaacgcgg gcctggcccc cggcgacgtg gacgcggtcg
46601 agggccacgg caccggcacc gtcctcggcg accccatcga ggcccaggcg
46651 ctgctcgcca cctacggcca ggaccggccc gccgaccggc cggttggtgct
46701 cggctcggtg aagtccaaca tcggccacac ccaggccgcc gccggcgccc
10 46751 ccggcctgat gaagatggtg ctggccctcc aacacggcac gctgccgcgc
46801 accctgcacg tcaccgagcc ctcgaccgg gtcgactggt cggccggcg
46851 ggtgcggctg ctaccgagc ggaccgtctg gccgaggacg gatcgctcgc
46901 gtcgggcccgg ggtctcctcg ttcggcatca gcggcaccaa cggccacgtc
46951 atcctggaac agccgcccgc cgagcccacc cccacggccc ctgccgaccg
15 47001 ccccacccgg acgcccgcgg tcctcccatg ggtcgtctcg gcccgatcgg
47051 ccaccgcgct cgacgcgcag ctcgcgcgac tgcgggcggt cgccgccgag
47101 cgcccggaac tgccgcccgc cgacgtcgcc cactcgctcg tcaccagccg
47151 cgccaccttc gaacaccggg cggtcctgct ggccgcgccc gacggcatca
47201 ccgcgccgcg ccgcgccgag gcccgcgaa gcagcaccgc gttcctcttc
20 47251 tcggggcagg gtgctcagcg gtcggggatg gggcgtgaac tgcatgctgc
47301 tttcccggtg ttcgcgccgg cgttcgacga ggtggtggcg gtgttgatg
47351 cgaggttggc gacgggttcc ggtgggggtg tgcgctgctg ggaggtgatg
47401 tggggcgggg ggtcggagtt gttggatcgg acgcgtttca cgcagccggc
47451 gttgttcgcg gtggaggtgg cgttggtccg tttggtggcc tcgtgggggg
25 47501 tggggcctga gttcgtggcg gggcattcgg tgggtgagat tgcggcgcg
47551 tatgtggccg ggggtgttct gttggtggat gcgtgtcgtt tgggtggtgc
47601 gcgggcttcg ttgatggatg cgttgccggt gggtagcggt atggttgcg
47651 tggaggcggc cgaggcgag gtggtgccgc tgttggtcga tggggtggcg
47701 atcgccgcgg tcaacgggcc ggtttcgggt gtggtctccg gtgtggaggc
30 47751 ggccgttggg caggctgtgg atcagttggt ggagcggggc cggcggttcc
47801 gtcggttggc ggtcagtcac gctttccact cgccgttgat ggatccgatg
47851 ttgatgcct tccgggccgt cgccagggc ctggagtacc accagccgcg
47901 catccccgtg gtgtccaacg tgacggcgga ggtggccgcg gcggaggagc
47951 tgtgcgcggc cgactactgg gtgcggcacg tccgggcgac cgtgcggttc
35 48001 gccgacggcg tccgcaccct ggccgagcgc gccgccaccg ccttcctgga
48051 gatcggcccc gacggcgtag tgtccgccct ggccggggcc tgectgttcg
48101 acacggacgc cgaagtgggt cccgcgctgc gcaaggggag ccccgaggag

48151 cacaccgccc tcaccgccgc cgcccaactc cacgtggccg gcgtggacat
48201 cgactggacc gcggtcctgg cgggcaccgg cgggaggcgg atcgccctgc
48251 ccacctatgc cttccagcgc gagcgggtact ggccctcgct cgccgcacag
48301 gccccggcg acgccggcgg gctcggcctg gaagccgggc ggcaccgcct
5 48351 gctcggggcc gcgaccaccg tcgccggatc cgcggagatc ctgctcaccg
48401 gccgcctgtc caccaccgcc cagccgtggc tcgcgggtcta cgaggcggac
48451 ggccgcaccg tcctgccggc cgcggtcctc gccgaactcg ccgtccgcgc
48501 cggcgaccag gccgactgcc cgaccgtggc ggaactgacc gtgcgcgcac
48551 cgctcgtcct caccggcgcg gcggcccagc gcctccaggt ccgggtggcc
10 48601 gccccgacg acaccggccg gcgcgcgctg tccgtgcacg cccgaccga
48651 cgactcccc gacagcccc ctggacgtgca cgccaccgcg gtccctaccc
48701 acgacacccc gcagcccccg gcgccggaca ccggtggcc gccggagcgc
48751 gccgtgccgc tcgacgccct gccaccgcc accggcccgg cccggatcgc
48801 ggcggcctgg cagtggggcg acgaactctg cgcgagatc gaactccccg
15 48851 aaccggccc ggcgagcgg gcattcgccc tgcaccggc gctgctggac
48901 accgcggtcc gcgccggcgg cctgctggac ggcgacgcca ccctggacgc
48951 cctcggtcgg cggggcctcg ccctgcacgc cgcgtccgcc accgccctgc
49001 gggctccgct caccgggac ggacgggaca cctgggtctt ggaggccacc
49051 gaccgcagc gcgctccggt cgtctccgtc accgggtcct ccctgggcac
20 49101 gccaccgtc gaccggtcgg gggccggggc ggccgatgac ggcgcgaccc
49151 tgctcgacct ggagtgggtg cccgcgcgc aggccgcgc caccggcggc
49201 gaccacctcc cgtacgccgt gctcggcgat caactcgcgg agctggacgg
49251 gcagttgagg atcgcgggcg acgggcccgg gcgcgtcgca tcgctggccg
49301 cgctgctgga cggcggtgcg ccgctgcccc ggctcgtcct cgcgcgggtg
25 49351 ctgggcgtgc cgaccgggga aggcgacctg cccgccggcg tgcgcggcac
49401 caccacggcg gtgctggagc tgctgcagcg ctggaccgcc gacgcccgca
49451 ccgccgacag ccacctggtg atcgtcacc gcggcgccgt cgccgcggg
49501 gcggaggacg tgcacgacct ggcggcggcc ccggtctggg gcctggctcg
49551 ctcggcacag tccgaacacc ccggcagctt cctgctgctc gacctcgacc
30 49601 ccgccgatcc cgcgggagcc tcccgcgcg ccgcgccggc caccctggcg
49651 gccctgctcg acgcgggcca gaccaggcc gcggtgcgcg ccgacacgct
49701 caccgtcgcc cggtgaccc gggccggcga cggacccgag gccaccggcg
49751 gacaccgggt gcgggactgg gaccgcgacg gcaccgtcct gatcaccggc
49801 ggcaccggcg gcctggggcg cctcctggcc cgccacctgg tcaccggaca
35 49851 cggcatcaag cacctgctgc tcgccggcg ccgcggccc gacgccccg
49901 gcgcgcgggc cctgcgcgac gaactggccg ccctcggcgc cgaggtgacc
49951 gtcgccgct gcgacgtggc cgaccgtgcc gactggacc gactcctcgc

50001 gcaactgccg ccggagcacc cgctgaccgc cgctcgtgcac accgccggcg
50051 tcctcgacga cgccaccgtc ggcaccctga cgcccagagc gctggacacc
50101 gtctcgcgcg ccaaggcgga cgccgcctgg cacctgcacg acgccaccgc
50151 cgaccgcgac ctggcagggg tcgtgctgta ctctcggtc gccggtgtca
5 50201 ccggcggccc cggccagggc aactacgccg ccggcaacac gttcctcgac
50251 gcgctcgccg cgcaccgcgc cgcccagggc ctgcccggac tgtcgtggtg
50301 ctggggaccg tgggggcagg acgccggcat gaccggcacc ctcggcgccg
50351 ccgacctggc ccgcttgagg cgctccggca tgccgccgt caccgccgaa
50401 cagggcctgg ccctgttcga cgccgccggc gcccgcgcg acgggttcgc
10 50451 ggtggcggtg cggctcgccc gtggcgccgc cgcaccgggc gccgacgagg
50501 tccccgcggg gctgcgtgcc ctggtgcgcg gccggcgccg caccggcgcc
50551 gcggccgggc acgccggtgt actggcccgc cggctggccg ccctggacgc
50601 cgagcagcgg catcaggcgc tgctcgacct ggtccgcacc gagacggccg
50651 cgggtgctcg ccactccggg gcggacgccg tcccgccga gcgggacttc
15 50701 aaccggctgg gcttcgactc gctgatggcg gtcgaactgc ggacgcggct
50751 ggccaccgcc accggagccc ggctgccggc cagctcgtc ttcgaccacc
50801 cgacgccgga cgcggtcgcc cggcacctcg cgtcgacgct gcccggtggg
50851 acccgggccg gtccggaccg tccccgctg gccgaactcg accggatcgc
50901 cgccgagttg tcgccggagg gcgcggacga cgccaccga caggcgctcg
20 50951 tcgggcggtg gcggcacctg ctggcgaggt gggacggcac ccgacaggac
51001 ggcggtggga cgaccgtcga cgaccgcac gaagcggcga gcgccgaaga
51051 ggtcctcgcc ttcacgacc acgagctcgg ccggcaggcg gactcctgac
51101 ccgccccact ccgctcgctc gcgcgcacca catctgagga aggtttcacg
51151 gaccatgccg gacgaaaaga agctcgtcga ctatctgaag tgggtcacga
25 51201 aggacctcca ccagaccgc cagcgccttc aggaggtgga ggccggcgcc
51251 cacgaaccgc tggcgatcgt cggcatggcc tgccgcttc ccggcggtgt
51301 gcgctccccg gaggacctgt gggagctgct gtccgcgggc cgggacggca
51351 tcgggcccgt ccccgccgac cgcggctggg acctggcggc gctggccggc
51401 gacgggcccg gtcgcagcgc caccaggaa ggcgggttc tgcccacgc
30 51451 ggccgccttc gaccggggt tcttcgacat ctccccgcgc gaggcgctcg
51501 ccatggaccc gcagcagcgg ctgctgctgg agaccgcctg ggaggccgtc
51551 gaacgctccg gcatcgacc ggccgggctg cgcggcagcc gcaccggcgt
51601 tttcgtcggc accaacggcc aggactacgc gcacctggtc ctgcgcgcgc
51651 aggacgacat gggcggttac gcgggcaacg gcctggccgc cagcgtgctc
35 51701 tccggccgac tggccttcgc gctcggcctg gaaggcccgc ccgtcacctc
51751 cgacaccgcc tgctcctcgt cactggtgac cctgcacctg gccgcacagg
51801 ccgtgcgcgc cggcgaatgc ggcctcgccc tggccggtg cgtcacggtc

51851 atgacgacct cgtcgagctt cgccggcttc agcctccagg gcggcctggc
51901 gccggacggc cgctgcaagg cgttcgccga ggcggccgac ggcaccggct
51951 ggtccgaggg catcggcctg cttctcgtcg agcggctctc cgacgcgcag
52001 cgcaacggcc acccgggtgct cgccgtgctg cgcggtccg ccgtcaacca
5 52051 ggacggcgcg tccaacggcc tcagcgcgcc caacgggtccg tcccagcagc
52101 gggatcatccg ccaggcgctg gccggcgccg gactcgctcc cggcgacgtg
52151 gacgcggctg aggcgcacgg caccggcacc cggtcggcg accccatcga
52201 ggccgggtcg ctgctcgcca cctacggcca ggaccggccc gccgaccggc
52251 cgttgtggct cggtcgggtg aagtccaacc tcggccacac ccaggccgcc
10 52301 gcgggcgtcg ccggcgctcat caagatggtg ctggccctgc ggcattggcgt
52351 cctcccgag accctgcacg tggacgcgcc ctctcgcac gtcgactggg
52401 agagcggcgc ggtgcggctg ctcaccgcac ccgtcgctg gtccgagggc
52451 gacgaccggg tgcgcgggc cggcgtctcg tcgttcggca tcagcggcac
52501 caacgcccac gtcattctcg aacaagcccc cgatcagccg gaaccgaccg
15 52551 cggaagagac ggctgccgcg gcgcccggcg gcaccgccga ggagcgggccc
52601 gccgctcccg tcgccccgcg cgccgtgccg tggccggctg cggcacgcac
52651 cgccggcgcc ctcgacgccc aactgggtccg ggtccgcgcg ctgaccaccg
52701 cgcccgcccg caccgcgcc gacgtcggtc acgcgtggc caccgccgt
52751 acccccttcg agcaccgggc gctgctggtc cagcaggcg gcgccgtcac
20 52801 cgaggtggcg cgcggcgcg tccccaccgg tgaccggggc gggctggccg
52851 tgctgttctc cggacagggc tcccaacggc cgggcatggg gcgcgaactc
52901 cagccccgt acccgtctt cgcccgccg ttcgacgaga ccgtcgccct
52951 gctcgacgcc cggtcggca cgtcgctgcg cgacatcgtc tgggaccagg
53001 accgcacccg gctcgacgac acccgccaca cccagccgc gctgttcgcc
25 53051 gtcgaggtcg cgctgtaccg cctgctggcc tcctggggca tccggcccga
53101 ccacgtcacc ggacactcca tcggcgagat caccgcggcg cacgtcgccg
53151 gtgtgctgac cctcgcgga gctgcaccc tggtgccgc ccgcgccacc
53201 gccatgagcg aactgccgcc cggcggcgcc atggtggcg tggaggccac
53251 cgaggacgag gtgcgtccgc tgctcaccga cgacctcgcg atcgccgcg
30 53301 tcaacgcccc ccggtccgtg gtcgtcgccg gcgccagga cgccgccctc
53351 gccgtccgcc ggcacttcga cgacctgggc cgccggacca cccggctccc
53401 ggtcagccac gccttcact cgccgctcat ggaccgatg ctcgacgcct
53451 tccggacggc cctcgccccg ctgaccttcg ccgagccgga gatcccggtc
53501 gtctccaacc tcaccggcct cccggccacc gccgaggaac tcgccacccc
35 53551 gcaactactg gtgtgccacg tccggcaggc cgtccgcttc ggcgacggcg
53601 tgcgcgccct cgccgaccgc ggcgtcgga cttcctcga actcgcccg
53651 gacggcgtgc tgtccgccct ggtccgggag aacctcccc agccgggcct

53701 ggtcgccgtg cccgtgctgc gcaaggagcg gcccgaggag accaccgtgc
53751 tggccgccct gggaaccctg tgggcgcacg gcgcggacgt ggactgggac
53801 gcggtgttcg ccggcacccg caccocgcag gccgaccocg tcgagctgcc
53851 gacgtacgcc ttccaacgcg cccgtactg gccaccctc ggcgcccgcc
5 53901 acggcgaccc ggccgacctc gggcagaccg ccgcgcacca cccgtgctg
53951 ggcgccgcg tcaccctcgc cgacgcgcgac gagaccgtgc tcaccggccg
54001 cctcgcgctg ccctcccacc cctggctcgg cgaccaccgc agcgacggcc
54051 ggatcacctg ccccgcgctc gccttcgccg aactcgccgt ccgcgcggc
54101 gacctgagcg gcaccccgca cctggcgcg ctcgacctgc cggcgccgt
10 54151 caccctcggc gacggcgaca ccgtcaccct ccaggtecgg gtcggcgccc
54201 ccgaccccg cggccaccgg ccgtgaccg tccacgccg cctcgagcc
54251 accgaggacg ccccttgac cactgcgcg accggtctgc tcgccccgga
54301 cgccccgaa gcgcccgcg atccgatcgg cccggccgac gccgggtggc
54351 cgccgcggga cgcccgccc gtgcccgtcg ccgacctga cgcgccggcc
15 54401 accgcgcag gccgccacta cgcccccat ttccagggcc tgaccgggt
54451 ctggcggcgc gacggcgagg tcttcgccga ggtggccctg cccaccgcca
54501 ccgcgcgca ccgcgcctc ggcatccacc ccgcgtgct ggccaccgcg
54551 ctccgcgcca ccgcgcact ggacgacgac cacaccgcc gccacacccc
54601 cgaaccgacc ggcatcaccg gactcgccct gcacgccacc ggggccaccg
20 54651 cactcggggt ccggctcacc gcgaccgggc ccgacaccgt ggccctcgcc
54701 gccgcggacg ccacgggcg cgcggtcctg accgcgaca ccgtcaccct
54751 cggtccccg caggaccgcc cggtcccg cccggccggc cacaccgggc
54801 agggcgccct gttccacctc gactgggtgc cggtcgacc cggcagccga
54851 gccaccggca cccgtgggc cgtcgtcggc gacgacgaac tcgacctcg
25 54901 ctacgccctg caccgcgcc acgagacgg cagtgcctac gcggcgtcgc
54951 tgggcggagc catcggcgac agcggctctg gcgccgacgt cttcctcgtc
55001 cccgtcgtcg gcggcccgga cgccgggcc gacgcggtgc acgcgtcac
55051 cgcccgccc ctggggctgc tccaggagt gctgaacgag ccgcggttg
55101 ccggcgccc cctggtctc gtcaccgcg gcgcgctcg ggtgcccggc
30 55151 gagaccgtca ccgaccggc cggcgccgc gctggggcc tgctgcgtc
55201 cgccagacc gagaaccgg gcagtctgct gctggtcgac ctcgacgacg
55251 cgttcgggtc cgccgggatg ctgccgcac tctcaccct cgacgaacag
55301 cagctcgtcg tccgcgacca cgcggtccgc gccgccgcc tggccggct
55351 gccggagccg gccgcggca ccgcgcggc ccgcgcctg gaccggacg
35 55401 gcaccgtcct gatcaccgg gccaccggc gcctgggcg cgcgctcgcc
55451 cgccacctg tcaccgtcc cggcgccgc cactgctgc tcgcccggc
55501 ccgcggccc gaggcgcgg gcgccgga actggtggc gagctgaccg

55551 cacagggcgc ggacgtgcgg gtggccgcct gcgacgtcgg cgaccgcacc
 55601 gccctcgacg cgctcctcgc cacggtcccc gcggcgcacc cgctgaccgc
 55651 cgtcgtgcac accgccggcg tcctggacga cgccctgacg ggctcgctca
 55701 cccccgacca actggccacc gtgctacggc ccaaggccga cgccgcctgg
 5 55751 catctgcacg acgccaccgc cggcctcgac ctggccgggt tcgtcctgta
 55801 ctctcggtc tccggcgctc tgggcagccc cggccagggc aactacgccg
 55851 ccgccaacgc ctacctcgac gcgctcgccc ggcaccgcgc cgaccagggc
 55901 ctcccggcgc tctccctcgc ctggggcccc tggggtcggg gcagcggcat
 55951 gaccgcgtcg gtcagcgacg ccgacctgga gcggatggcg cgcgccggcc
 10 56001 tgccgccgct gaccgtcgag gacggcctgg ccctgttcga cgccgccgtc
 56051 ggccgccccg agccggccct ggtgcccgag cgcatcaacg tcgccggcct
 56101 gcgggaccag caggcactgc cggcactctg gcgcgacctg gtaccgcggg
 56151 cccgccgcac cgcgccacc gccgaccgtc ccccggtcac ggtgcgcgag
 56201 cgcctccgcc acctcgacga gaccggccag gagcagctgc tcatcgacct
 15 56251 cgtcgtcggc tacaccgccg gcctgctcgg ccaccccgac cccaccgccg
 56301 tcgatcccga acggggcttc ctggagctgg gcttcgactc cctgggtctcg
 56351 gtcggcctgc gcaaccagct cgccgagatc ctcggcctgc gcctgccctc
 56401 gtccatcgtc ttcgacagca agtcgccggt gaagctggcg cgttggtcgc
 56451 accaggaact cgccaacggc cccagccgg gcgccaccgg ccccgccgcc
 20 56501 gcggacgccc gtcccgcggt gcgctccgac gacaccctgg agggcctggt
 56551 ctacaacgcg gtgcgcggcg gcaagctcgt cgaggcgatg cggatgctca
 56601 aggcctcgc caacaccgcg ccgatgttcg acaccccgcc cgagctggag
 56651 gagctctccg agccggtgac gctcgccgac ggcccggggc ggcccgggt
 56701 gatcttcgtc agcgcgcccg gcgccaccgg cggcgtccac cagtacgcgc
 25 56751 gcatcgccgc gacttccgc ggcagccgcc atgtctccgc gctgcccctg
 56801 atgggcttcg ccccgggcga gctcctcccg gccaccagcg aggcgcggc
 56851 ccgtatcgtc gccgagagcg tcctgatggc cagcgagggc gaaccgttcg
 56901 tcattggtcgg ccaactccacc ggcggctcgc tggcctacct cgccgcggc
 56951 gtctggagg acacctggga cgtccggccc gaagcggtgg tcctcctcga
 30 57001 caccgcgtcc atccgtaca accccggcga gggcaacgac ctggaccgca
 57051 ccacgaggtt ctacctggcc gacatcgact cgccctcggt gacgtcaac
 57101 agcgcccgga tgtccgcat ggcccactgg ttcattggcg tgaccgacat
 57151 ccaggcgccc gcaccgaccg ccccccacct cctcgtgcgc gccgcccggg
 57201 ccctcgacgg cttccggctc gacacctcgt ccgtccccgc cgacgaggtc
 35 57251 cgggacatcg acgccacca cctctccctc gccaaggagc actcggcact
 57301 gaccgcgcag gccatcgagg gatggctcgc ggaactgccg gaccccgcg
 57351 cctgatctcc ggcccggccg gccccgaca ggccggggcc atccccccac

57401 cggggcggcg ccaccaggt gcgccccgc gcacccggcc ccgcgcaggc
57451 agccggagcg cccatcccc cgcatggcac caccctccag aggagtcctt
57501 ccatgagcac accgaccgca ccgccctccc tgaaagcgga ggtgccgccc
57551 gtcctgcgcc tgagcccgct gctgcgcgaa ctccagtccc gcgcccccg
5 57601 ctgcaaggtc cgcaccccc cgggcgacga gggctggctg gtgacccggc
57651 acaccgaact caagcagctg ctgcacgacg accggctggc ccgcgcccac
57701 gccgaccgg ccaacgcccc gcgttatgtg cacaacccgt tcctggacct
57751 gctcgtcgtc gacgacttcg acctggcccc cagctgcac gccgagatgc
57801 gtccttgtt caccgccag ttctcggccc gccgcgtcat ggacctgacg
10 57851 ccgaggggtg aagccctcgc cgagggggtg ctggccact tcgtcgcca
57901 gggaccgccc gccgacctgc acaacgactt ctgctgccg ttctccctgt
57951 cgggtgctgt cgcgctcatc ggcgtcccg ccgaggaaca ggggaagctg
58001 atcgccgccc tcacaaact gggcgaactc gacgaccgg cagcgtcca
58051 ggaaggccag gacgagctgt tcggcctgct gtccggcctg gcacgccgca
15 58101 agcgcacac acccgaggac gacgtcatct cccggctctg cctgaagggtg
58151 ccctccgacg agcgcacgg ccgcatcgcc tccggtctgc tcttcgccg
58201 cctggacagc gtcgccagcc acatcgacct gggcacggtg ctgttcaccc
58251 agcaccgga ccagctcgcc gcggccctgg ccgacgagaa gctgatgcgc
58301 ggcgccgctg aggagatcct gcggtccgcc aaggccggcg gttcgggtgt
20 58351 cccgcgttac gcgaccgcc atgtaccgat cggcgacgtg accatcaggg
58401 ccggcgacct ggtgctgctg gacttcaccc tgggtgaactt cgaccgcag
58451 gtcttcgacg agccggagct cttcgacatc cggcgcgccc ccaaccgca
58501 cctgacgttc ggccacggca tgtggcactg catcggcgcg ccgctggccc
58551 ggggtcaatct ggcaccgcc tacacctgc tgttcacccg cctgcccggc
25 58601 ctgcggtgtg tgcgcccggt cgaggaactg cgggtgctgt cggggcagtt
58651 gtcggccggc ctgacggagc tgcccgctac ctggtgacgt gatgtcggac
58701 ccgcccggcc cggttccggg caggcggaag tgagggcccc cgccccgctc
58751 ggccgggggc cctcacgcac gggggagcgg gtcctcagt ccgccgcgac
58801 gctgatcgcc ccggacgggc agagcgccgc cgcgtcgcgt acgtcgccgg
30 58851 gatccgcggc gtcgccgcc ccggccagga cggtcaccag accgtcgtcg
58901 tcctggtcga acaggtccgg tgcggtcagg acgactggc cggccccgac
58951 acagcggccg ggggtccagg tgatgcgcac gatcgttcct ccgagggctg
59001 gttcgtcgcc ggggtcgggc cgccgcccgg cggcccgacc gggctcactt
59051 ccaggtgacg ggcagttcgt gcaggccgaa cagcaccggc tcgtacttca
35 59101 acggcagttc ctccaccggc acggcgagcg tgagggaggg gatccgggcg
59151 aagagcgtgc ggtaggcgat gtccatctcc tcccgcacca ggttctggcc
59201 caggcactgg tggacaccgt agccgaacgc cagtgcgag cgcgccgagc

59251 gctcgggggtc gaactccgag ggcgccgcga aggccgtggc gtcgtggttg
 59301 gcggcggcga tcagcggcac gatgccctcg ccggcccga tcagctgacc
 59351 gccgatctcc acgtcctcga cggccaccgc gaatgccacc agatcggcca
 59401 ccgagtagta gcgcagcagt tcctcgacga tccggtcgtc gccgatccac
 5 59451 cgcgggttg tgagcaactg caccacgccc aggccgatgt tgttcgccgt
 59501 cgtctcgtgc ccggcgatca gcagcagcat cgccaccccg gacagctcct
 59551 gccgggacag tgtgcccgcg gcgatcagcc ggctgatcag gtcgtcaccg
 59601 ggtcgccgct gcttgatctg gatgagtcgg ccagatagc gcagcagcgc
 59651 ctgggtcgcc ttgtcgcgct cctcgtcggt cgaactgagc cggaccagca
 10 59701 cccgggtccg gtcctcgaag aagtcccgt cgacctggg caccgcgagc
 59751 agactggaga tcaccaggga cggcaccggc agcgcgagg actcgacgag
 59801 gtccgccgag gtgcccgcg ccagcatcgc gtcgatccgc tcgtccacga
 59851 tctgctggat caccggccgc agttccgcga ccttgcgac ggtgaactcg
 59901 gggatcaagg tcttgcgga ccggccgtgc tccggtgggt ccatcgccac
 15 59951 gaaccacccc ggcacctcgt actcgacgg ggccccaccg gtccgaccgg
 60001 ccttggggaa accgggtttg gaggggtcgg cgctgatccg ggggtcggtc
 60051 agcaccgccc ggacgtcctc gtgccgggtc accaaccaga ccgggcccgt
 60101 ggggagttcg gagcggaacca ggcccgcgcc gcccggtag gtcggtact
 60151 ccggcggggg gaagggccgg cccggcctgc gcagcgggaa ggccaccggg
 20 60201 cactgcggcg cggcggtccg cgcgtcggct tcgggtgctca tgccactccg
 60251 tagaacgcgc ggatccggcc ggtgatgaac tcctgctcct gtgcggtcag
 60301 tccggtgtgc gtcggcagat agaagccgtc ggcgctgagc cgggcccgt
 60351 tcaacgcggg ccagtcggcg gagaagtaac cgggctgacg gtcctcggc
 60401 ttgaagaaga cccgggtctc gatgccctcg ccggcgaggt acgcgcacag
 25 60451 ctcgtcccgc cgtcggccc gcaggtcgta catccacagc acgtcgcgcg
 60501 gcggcatcag cgtgatgccg gggatgtcgc gcagcgcctc gtcgtagcgc
 60551 ttctcgatgt cgcggcgag ggcgaggatg gtgtccagct gtcggtctg
 60601 cgcgagcgcc accgcggcct gcatgttgg catccggaag ttgtaggcca
 60651 gcttcttgtg caggaagctg tggctcttg tgaacgcat cggccgcaga
 30 60701 tgggccatct gtcggccag gtgcgggtcg tgggtcaggc agacgccgcc
 60751 ctgcgggcc gagatgatct tgttggcgaa gagcgagaaa caggcgatgt
 60801 cggccgcgg ccgcaccccg tgcgcctcgg cggagtctc caccaccgcg
 60851 aggttgact cgtacgccag gttcagcacg gcgtccatgt cgcactgccg
 60901 gccgtagatg tgcaccggca tgatcacttt ggtgcgggg gtgatcttct
 35 60951 cctcgatgcg ggacacgtcg atgttcaggt cgtcgccgca gtccacgaac
 61001 accggcgtgg caccggtgta ggtcaccgcc caggcgagc cgatcatcgt
 61051 gaactccggg acgatcacct cgtcaccggg gccgacgcc agcgcgcgca

61101 gcgccagcgt cagcgccgtg gtgccggagg agcaggcgac gccgaacggc
61151 acgtcgttgt acgcggcgaa cgctcctcg aaccgcctga cgtacggccc
61201 ctgcgaagag atccagccgc cgccgacggc ctccgtcaca tagtcgagct
61251 cgcggccctg gagccacggc atggacaccg gatacgtaaa ggacatgggt
5 61301 tttgagtctt cctcggtcag tcggttgcca gggcggaag gccgagcagc
61351 aggtccgccc cgcgggccc ggcgccggcg tcccgcagca gcccggcgaa
61401 gtgctccgcg cgctcggta aggagggttg gtcgagcacg cgggtgatct
61451 tgtccaggac gtcctcggtg tccacggctt ccggccggtc cagggtcagg
61501 ctacccccga agtcctggcc ccggatcgcc tggtcgtcgc agtccacca
10 61551 caacggcccg accaccagcg gctttccgaa gtacaggccc tcgtggtagc
61601 cgttgccacc ggcatgggtg aagaacgcct tcacgttcgg atgggccagc
61651 acgtccagct gcgacggcac ccagccctcg atccgcaggt tgtccggcag
61701 ctcgccggcc ggccggcagca actcctgttg gccgcgccc agtttccaca
61751 acacctggtg gcccggccc tccagtcgcc gggcgacctc caccagcgac
15 61801 gccacctgct cacgggtcag ccgggtgatc gtgccgaagc ccatgtacac
61851 cacggacttc tgcgccgaca gccagtccga caggccgtcg tcgtccgggtg
61901 cctggggcag cggcggcacc atcgtgccc ccagccgcag ctccggatgc
61951 atcgggaacg ggtagtccaa ctcccttacg gagtagcaca agacctgctc
62001 cgcatgggtg atccgcgcca tcactgccc cgccctgggc gcgatgccc
20 62051 gtcgggtgcg gaccgggtt tctcctcga cgacctgcg gacgtccgac
62101 gtcaggaaca tcccagcgt ccgcagccgg aacagctggt tctcgatccg
62151 ctgagccagg gacatcgcg ccggcagccc cgagtgcggc accgggaaac
62201 ccgacggggt gtaggacttg gcgaacggga cgtgcgagg gaggacgttg
62251 ctcggcacga acggcaccac gagcacgaac ggaatgccct tggatgcgc
25 62301 cagctcgta ccgaactggc acatgctctc gatcaccatc agcgccggct
62351 cgacctctc gacgatctcc tccaggcggc ggtacttcgc catccgcgac
62401 tccggcgca acgaatgcc aatcacgcc gctgcgccct tgaaccgca
62451 ccgctcgctc acctccgat acgtcgctc gtcccatgt accgccgaca
62501 tctgcgagac ggtgtcgcc agcgacgca accgaaccgg gctgccgtcc
30 62551 accacggccc ccacctgct gcgcgcttc tcgtcgggtg cgaaccacag
62601 gtccgccacg tcgcgcggg acaattcccc ggccagcac agcagcggat
62651 tgagcaggcc gctttcgca taactgacga acaggatcg ccgccgatc
62701 gcgcccattg acaacacccc tcggaatgt gcgggccgcc gggcccgcgc
62751 gccacgcacc cgcccggccc ggtcgccgg tgagtgcatt cgccgacgcc
35 62801 gccacccgag gcgcgtgtt ccggaaggaa gggtcaccgg ccggcaccgc
62851 gaacgcgcc cgtggaaaac gggtcggta cttggtctca tgccacggac
62901 cggggaatca ctagtcttc gcgcgcgac gccctttccc ggccgtgtgg

62951 ccaatgcccg tccccggcgc ccgtcattcc ttagggaaaa gtacagcggt
63001 tgcgaacgta cgatccggca cgcagagggtg acctgaggcc aacttttccg
63051 caggggtgag caaggcatga cgatcggagc cgacgaggac ccggtggtgg
63101 tcgtcggaat ggcctgccgt tatccgggtg gggtcgccgg cccggaggac
5 63151 ctgtgggaac tggtcgcac cggccgcgac gcgaccaccg ccttcccgga
63201 cgaccgcggc tgggacctgg ccgactggc cggcgacgga cccggccgca
63251 gcgcgaccgc cgaggcgga ttctcaccg gcgcccgcga cttcgacgcc
63301 gccttcttcg gcatgtcgcc ccgcgaggcc gtctccaccg acccgcaaca
63351 gcgcctcgtc ctggagaccg cctgggaagc cctggagcgc gccggcatcg
10 63401 acccgactc cctgcgcggc agccgcaccg ggtcttcgt cggcgccagc
63451 ggccaggact acgccgccgt caccacgcc tcgcccagc acctggacgg
63501 acacgccctc accggcctgg ccccgccgt cgctccggt cgctggcgt
63551 acgtcctggg cctcgaaggc cccgcgtca ccgtcgacac cacgtcctcc
63601 tcgtcgctgg tcgcgctgca ctggcggtc cgcgccctgc gcgcggggga
15 63651 gtgcagcacc gccctggccg gcggcgtcac ggtgatgtcc acccggccg
63701 ccttcgtcgg ccacaccga caggcgggcc tcgcgccga cggcgctgc
63751 aagccgttct ccgacgacgc cgacggcacc gcctgggcgg agggcgctcg
63801 catcgtcgtc ctggagcacc tgtccaccgc ccgcgccgc ggcaaccccg
63851 tcctcgccgt gctgcgcggc tcggccgtca accaggacgg cgctccgac
20 63901 ggctcaccg caccagcgg tccgcccag gaacgcgtca tccgcgccg
63951 cctcgccgac gccgactcg ccccgccga catcgatctc gtcgaggcgc
64001 acggcaccgg caccggctc ggcgaccccg tcgaggcccg ggcgtgctc
64051 gccgcctacg gccaggaccg ggaccggac cgaccgctgc gcctcggttc
64101 cctgaagtcc accctcgcc acgcacaggc cgccgcggc atcgcgagg
25 64151 tgatcaagac cgtcctgacc ctgcggcacg gcctgatgcc gcgcatccgg
64201 cacctggcca ccccccaccg ccaagtcgac tggteccagg gcgcctggc
64251 cccctcacc gaccacacgc cctggccacc ggccgaccga ccgcgccgcg
64301 ccggcgctc ctccttcggc atcagcggca ccaacgcca tgtgatctc
64351 gaagaggcgc cgcccgccga cgtccccgtc acccggccg gcaccctcg
30 64401 cccagcacc gtcccctggc cgtctccgc cgccacgcc gaagccctcg
64451 acgcccact cgcccggtc cgcgccacc tgcgcacca ctggacctg
64501 gaccgctgg acgtcggtc ctcctggcc accggccgc ccgcgtccg
64551 ccaccgggcg gtcctcctgc cgcccgccga cggcaccgcc gcggacgcg
64601 tcgagcacgc ccgcggtcg gccaccagc gccgcaccgc cgtcctctc
35 64651 tccggccagg gcagccagc cccgggcatg ggccgcgaac tcgccgccg
64701 cttcccggtg ttcgccgacg cactggacga cgcgtgcgc gccctggacc
64751 ggcacctgga cggcccggtg cgcgaggtga tgtggggcac cgacgccgcg

	64801	ctcctggacc	ggaccggctg	gacccagccc	gccctgttcg	ccgtcgaggt
	64851	cgccctccac	cgctggctcg	cgccctcgg	cgtaaccccc	gacttcgtcg
	64901	gcggccactc	cgtcggcgag	atcgccgccg	cccacgtcgc	cggcgtcctg
	64951	agcctggagg	acgcctgccg	cctggtggcc	gcccgcgcca	cgctgatgca
5	65001	ggcgctcccg	gcccggggcg	cgatggccgc	gctggaggcc	accgaggacg
	65051	aagtggcccc	gctgctcggc	gcacacctcg	cgctggccgc	cgtaaacggc
	65101	cccaccggcg	tcgtcgtcgc	cggagccgag	gacgccgtgc	ggcaactgac
	65151	cgcccgcttc	gcccagccgc	gcccggcgac	cagccggctg	gccgtctcgc
	65201	acgccttcca	ctcgccgctg	atggagccca	tgctcgacgc	cttcggggac
10	65251	gtcgtgagcc	gactgacctt	ccaccagccg	tcgatcccg	tggtctccaa
	65301	cctcaccggt	gaactcgccg	gcagtgagat	caccagccgc	gagtactggg
	65351	tccggcacgt	ccgcgacacc	gtccgcttcg	ccgacggcat	caccgcactg
	65401	gccaaggccg	gcgcccagct	cctgatcgaa	ctcggccccc	gcggcggtgt
	65451	gtccgcgatg	gcccgcgaca	ccctcgcccc	cgacagcacc	accgacgtcg
15	65501	tccccgccct	gagcaaggga	cggcccggag	agaccgcctt	cgccggcgcc
	65551	ctcgcccgcc	tgcaaccctt	cggcgtcccc	gtcgactggc	ccgccttcta
	65601	cgccggcacc	ggcgcccggc	gcgtcgaaat	gcccacctac	gccttccagc
	65651	acgtgcgcca	ctggcccacc	ccgccccggc	cgaacggcgc	cgggcccggc
	65701	gccctcggcc	accccttgct	cggctccggc	gtcgaaactg	ccgacggcgg
20	65751	cggcaccgtc	tgctccggcg	ccctctccct	ccgcacccac	ccctggctcg
	65801	ccgaccacac	cgtcgcccgg	cgggtcgtgc	tgccggccac	cgcgctgctg
	65851	gaactcgccg	tgcgcgccgg	cgacgaggcg	ggctgcgacg	tcctgcacga
	65901	actccacctc	accaccccg	cggccctg	cgacgacgcc	gccctgcacg
	65951	tccaggtgca	cgtcggcccc	gcccagacca	ccgggcggcg	cgccgtcacc
25	66001	gtccacaccc	gcccagacca	ccaccggcc	ggcgactgga	cccgatgcgc
	66051	caccggcacc	ctcggcagca	ccccgccgtc	cgacgcccga	gccgccacgg
	66101	gcggcacccc	ggccgcctgg	ccgcccggcg	acgccgaacc	cctcgacctc
	66151	gccgaccact	acgaacggct	cgccgaccgc	ggcttcgact	acggcccgcg
	66201	cttcgcgggc	ctcgggggcg	cctggcgacg	cggcgcgag	atcttcggcg
30	66251	acgtggaatg	cccggccggc	accgcccagc	acgccccga	ccacggactg
	66301	caccccgccc	tgctcgacgc	ggcccggcac	gccgccatgg	cggtggacgg
	66351	caccgtgccc	gtcgccctgg	acggcgtccg	gctgcacgcc	gtcggcgcca
	66401	ccgcgctg	ggtccgcac	cgccccacca	cgaccggcac	gctgacctc
	66451	accgcggtcg	acgtgcacgg	cgcgccggtc	gtcaccgtcg	aggccctcac
35	66501	cgcccgcccc	ctgaccgacg	aggaacgcgc	cgcccccgcg	acgcggcggc
	66551	aggcccgcg	cgagacgccc	gccgacgccc	gcccggcccc	gcccggggcg
	66601	gcccggcccc	gcccggcgcg	cgaaccctc	ccggacacca	ccgggtccca

66651 cccacccgcc ggccacctcg ccgcgctgcc gccggccgcc cgggagcgcc
66701 agctgctgga cctggtgcgc acccaggccg ccgccgtcct gggccacccc
66751 ggccccgagg ccgtcggcac ccgcagcgtc ttcaaggagc tgggcttcga
66801 ctcgttggcc ggctcgaac tcgccgaccg gtcaccgcc cgcaccggac
5 66851 tgcgcctgcc ggccaccctc gtcttcaact tccccacccc cgaacgtgct
66901 gccacccgcc tcggggaact cctcgcgcga accgcccccc tcgaccccgg
66951 ggcgtaggga gaggaactca ccaggttcga ggcgatcgtg acgaacctgc
67001 cgaggacgg ccccgaaacgc cgggccgctc cggaccggtt ggacgccatc
67051 gtctccgcac tccgccagaa ctgcctgca gaggtgccct cctcggacga
10 67101 ggacatcgac acggtgtcgg tcgacagact gtcgacatc atcgatgaag
67151 agttcgaaac cacatagaga aattgttgtt ttcgttcgcg acccgatgac
67201 gaggacggac cgatgcagga accccagcaa ggccagccgg accagcagga
67251 gaaaatcgtc gactatctcc ggcggtcac ttcagatctt cgccgtgccc
67301 gccgccgat tggcgaactg gaatccaagg acaacgagcc catcgccatc
15 67351 gtcggaatgg gctgccgact tcccggcggc gtcaattcgc cggaatccct
67401 gtgggacctg gtgcgttcg gcggcgacgc catttcgga tccccgtcg
67451 accgcggctg ggacctgaa accctcaccg gaaacggcga cggcagcagc
67501 gccaccacg aaggcggatt cctctacgac gccgcggaat tcgacgccgc
67551 cttcttcggc atctcgcgc gcgaggcgac tgctatggac cccagcagc
20 67601 gcctcctcct cgaagtcgcc tgggaggcgc tggagcgcgc cggcatcgcc
67651 cccacagccc tcgcggcag ccggtccggc gtgttcgtcg gtcctacca
67701 ctggggcgcg ccctcggccg acgccgccac cgaactgcac ggccacgccc
67751 tgaccggcac cgcgcgacg gtgctgtccg gccgcctggc ctacaccctc
67801 ggctcgaag gcccgccgt caccgtcgac accgcctgct cctcctcct
25 67851 cgtcgccctg cacctggcgg ccagtcctc gcgcgtcggc gaatcctcg
67901 tcgcgtgat cggcggcgtc acgatcctca ccgagccgtc cgtcttcgtc
67951 gagttcagcg ccagggcgg cctggcaccg gacggccgt gcaaggcgtt
68001 ctccgacgcc gccgacggca ccggttgggc cgaggcgctc ggtgtcctcg
68051 tcgccgagcg gctctccgac gcgcagcga acggccatcc ggtcctcgcc
30 68101 gtgctgcgcg gctccgccgt caaccaggac ggccctcca acggcctgac
68151 cgcccccaac ggccctccc aggaacgggt catccagcag gccctcgccc
68201 ggaccggcct gacccccgcc gacatcgacg ccgtcgaggc gcacggcacc
68251 ggcaccggc tcggcgaccc catcgaggcc caggccctgc tcgccacct
68301 cggccaggga cacacccccg accagccgct gtggctcggc tccctgaagt
35 68351 ccaacatcgg gcacaccag gcggccgccg gcgtcgccgg tgtcatcaag
68401 atggatcatg cgctgcgcca cgccacctg ccgccgaccc tcgacgccga
68451 cgcgccctec tcgcacgtg actggtccgc cggatcggtg cgcctgctga

68501 ccgagggcca gcagtggccg gagaccggac gtccgcgccg ggccgcggtg
68551 tcctcgttcg gcatcagcgg caccaacgcg cagccctgc tggaaacaggc
68601 accccacccc gcggacaccg cggacgcgg cgacgacgcc gcgcccaccg
68651 aaccggcccg cgcgcccgcc gcgctgccct ggatcgcttc cggacactcc
5 68701 ccgcaggcgc tgcgcgacca ggccgcgcc ctggccgcca gggtcgagac
68751 cgaccccgcg ctccgcccc aggacatcgg gcacaccctg cacaccgccc
68801 gcgccttgct cgaacgacgc gccgtcgtcg tcgccccga ccgcgccgaa
68851 ctctcgcgcg ctaccacga gttggccgcc ggccgggtccg cgaacgccgt
68901 cgtcgagggc ctgcgggacg tcgagggtcg gacgggtgtc gtgttccccg
10 68951 gtcaggggtc gcagtgggtg gggatggggg cccaactcct cgatgagtcg
69001 gcggtgttcg cggagcggat tgcgagtggt gcggcgccac tcgccgagtt
69051 caccgactgg tcgctggtcg atgtgctgcg ggggtgtggtg ggtgcgccgt
69101 cgttggagcg ggtcgatgtg gtgcagccgg cgtcgttcgc ggtgatggtg
69151 tcgttggtcg cgttgtgggg ttcccggtgt gtgttgccgg atgcggtggt
15 69201 ggggcattcg cagggtgaga tcgctgccgc ggtggtgtcg ggtgcgctgt
69251 cgttgcggga cggggcgcg gtggtggcg tcgggagtcg ggccattggt
69301 cgtgcgttg cggggcgggg cgggatgatg tccgtcgcg tgctcgggtga
69351 cgtgctcgaa ccgcggttg tcgagttcga gggcggggtg tcggtggccg
69401 ccgtcaacgg ccgcgctcc gtcgtggtcg ccggcgagcc cgaggcgtg
20 69451 gacgcgctgc acgcccggct gaccgccgac gacatccggg cccgccggat
69501 cgcggtggac tacgcctcgc actcgacca ggtcgaggac ctgcacgagg
69551 aactgctgga ggtgctggcg gagctggcg cgcgcacgtc ggaggtgccg
69601 ttcttctcga ccgtgaccgg cgactggctg gacaccgcgc ggatggacgc
69651 cggctactgg ttccgcaacc tgcgcggacg ggtgcggttc gcggacgcgg
25 69701 tggcggacct gctggcgcg gagtaccgcg cgttcgtcga ggtcagctcg
69751 cacccggtgc tgacgatggc ggtcttgac ctgatcgagg aggcgggggt
69801 caccggcgtc gcgaccgcca ccctgcgcg tgaccagggt ggcgcgggcc
69851 gcttctgct gtcggccgcc gaggtcttcg tgcgcggtgt ggacgtggac
69901 tgggcggggg cgttcgaggg gaccgggtcg gcccggtcg acctgccac
30 69951 ctacgccttc cagcgcgagc ggtactgaa caccgcacc gccgccgacc
70001 gcaccccggc cgacgcccc atggacgcc aattctggg cgccgtcgaa
70051 caggcgacg tctccgcgt gaccgccgcg ctggcaccg acgaggactc
70101 cgtcgccgcc atcctgccc gcctcacctc ctggcgccgg gcccgctccc
70151 agcgcaccac cctcgactcc tggcgtacc gcgtcacctg gacgccctc
35 70201 gccaggtgc ccgcgccac cctgaccggc acctggctgc tggtcaccac
70251 cgacggcatc gacgacacc atgtggcagg ggcgttgag agctacggcg
70301 ccgaggtgcg ccggtggtc ctggacgagg agtgcaccga ccgcgccgtc

70351 ctgcgggagc ggctggccgg cgcggaggac gtgaccggca tcgtctccgt
70401 cctcgccgcc gccgaggacg acgccgcacg ccaccccggc ctcacccggg
70451 gactcgcgct caccgtctcc ctctgtccagg ccctgggcca cgcgaggcg
70501 accgcgccgc tgtggttcct gaccgcggc gccttcgcca ccggcccgtc
5 70551 cgaccccgtc acccggcccc tgcagagcca gatcgcgggc gtcggctgga
70601 ccaccgcgct ggagcaccgg cagcgctggg gcggcaccgt ggacctgccc
70651 gacaccctcg acgcccgggc cggccagcgg ctcgccggcc cgctgtccgg
70701 cgccctcggc gccgaggacc agctcgccgt ccgcgcccgc ggggtactgg
70751 cccgccgcat cgtgcgtgcc ggacaccggc ccggacgacc ggcacggacc
10 70801 tgggcgccgc gcggcaccac cctgatcacc ggcggctccg gcaccctcgc
70851 cccgcagctc gcccgctggc tggccgaacg cgcgcgccg cactgtgtgc
70901 tggtcagccg gcgcggtgcc gacgccccg gagcgcccga actcatcgcg
70951 gaggcagccg agtcgggcac cgaggtgacc gtcgccgcct gcgacatcac
71001 cgaccgcgac gcggtcgccg cgctgctggc cgacctcac gccgacggcc
15 71051 gcaccctgcg caccgtcatc cagcccgccg ccgccatcga gctgtcccg
71101 ctgcgccaca ccaccgtggc ggagtccgcc gacgtcgtgc acgccaaggt
71151 caccggcgca cggatcctcg acgaactgct cgacgacgcg gaactggacg
71201 acttcgtcct gtactcctcc accgccggca tgtggggcag cggcgtgcac
71251 gccgcctacg tcgccggcaa cgcctatctg tccgcgctcg ccgagcagcg
20 71301 ccgcgcccgc ggactgcgca ccacctccat ccaactgggc aagtggcccg
71351 acgaccgggc acgcgagctg gccgaccgc accggatccg ccgcagcggc
71401 ctggagtacc tcgaccccca gctggcgctc accgcgctcc agcacgtcct
71451 ggacgacgac gagaccgtca tcggcctcat ggacatcgac tgggacacct
71501 accacgacgt gttcaccgcg ggccggcccc cgcacctctt cgaccagatc
25 71551 cccgaggtgc ggcgcgcct cgaccaggca tccgtcccgg accccgcggg
71601 cccggccgcc gacggcctgg ccgccgcct gcacggcctc gccgcgcgcg
71651 aacaggaccg gctgctgctc accctggtcc gcaccgaggc cgcgcgcgtc
71701 ctcggccacg cctcggccga gtccttcccc gagcgccgcg ccttcctgta
71751 cctcggttc gactcggtca ccgccgtgga cctgcgcaac cggctcgtgg
30 71801 ccggcaccgg actgcggctg ccctcgacga tggctctcga ccacccaac
71851 tgcgcggcgc tcgccgcgtt cctgaagacg acggcgctcg gcgtccccgg
71901 cgccgcaccg cagcagcacg ccgctaccgg caccgccggc gacgacgacc
71951 cgatcgccgt gatcgcatg agctgcgcgt accccggcgg cgcgcgcc
72001 cccgaggaac tgctgcggt cgccctcgac ggcgcgcgac tcattctcga
35 72051 gttccccgcg gaccgcggct gggacgccc gggcctgtac gaccgggacc
72101 ccgaccgccc cgccacacc tactccgtcc agggcggtt cctccacgag
72151 gccgcggct tcgatcccg cttcttcggg atctccccgc gcgaggcggc

72201 cgccatggac ccgcagcagc ggctcctgct ggagacctcc tgggaggcgt
 72251 tcgaacgcgc cggatcgcac cccgcgtcac tgcgcggcag cgccgccggc
 72301 accttcttcg gcgccagcta ccaggactac tcctccaccg tgcagaacgg
 72351 cacgggggag tccgaggcgc acatggtgac cggcaccgcg gccagtgtcc
 5 72401 tgtccggccg ggtctcctac ctgctcggcc tggagggccc cgcggtcacc
 72451 gtggacaccg cctgctcctc ctacttggtc gccctgcacc tggcctgcc
 72501 gtccctgcgc gacggcgaga gctccctcgc gctggccggc ggtgcccgcg
 72551 tgatggccac cccgcacgcg ttcgtcggct tcagccggca gcgtgccctg
 72601 gccaaaggac gccgctgcaa gccgttctcc gacaccgccg acggcatgac
 10 72651 gctcgcgcag ggcgtcggcg tcgtcctgct ggagcgctg tcccacgccc
 72701 gcgccaacgg gcaccgggta ctggccgtga tccgcggttc cgccgtcaac
 72751 caggacggcg cctccaacgg cctgaccgcg cccaacggcc cgtcccagca
 72801 gcgcgtcatc cgccaggcgc tcgccaacgc cggcctgacc ggcgcgcagc
 72851 tcgacgcggc cgaggcgcac ggcaccggca ccaagctggg cgaccccatc
 15 72901 gaggcccagg ccctgctcgc cacctacggc caggaccgcg acgccgaacg
 72951 gccgctgctg ctgggctcgg tgaagtcaa catcgccac acccaggccg
 73001 cggccggcgt cgccggcgtc atcaagatgg tgctggccat ggacgccggc
 73051 gaactgcccg gcacctgca cctcgacgcg ccctccagcc acgtcgactg
 73101 gaccgccggc gccgtggaac tgctgcgcgg gcgcaccccg tggcccagga
 20 73151 gcggggcggc ccgccgggccc ggtgtctcct cgcttcggcat cagcggcacc
 73201 aacgcccacc tgatcctcga acaggccccg gccaccgagc cgccagccga
 73251 ccccgaccgc ctccgggaca ccgccaccga caccgtcgtc ccctggccgc
 73301 tcgcccgaat gtccccggcc gccctgcgcg cccaggccgc ccggctcctc
 73351 gccaccgtcg agcacgaccc cgacctcccg cccgcccccg tgggccaacg
 25 73401 cctggccacc acccgccgcg ccctcgaaca ccgcgccgtc gtcgtcggcg
 73451 agcgccgcga ggacttctcg cgcgccctgg ccgccctgtc caccggcgcc
 73501 tcgacggccg gcctggtcag cggcatcgcc ggccccgacc ccgagggagc
 73551 ggtcttcgtc tccccggcc agggatccca gtggtgggga atgggcccgc
 73601 aactcctcgc cacgtccgag gtgttcgcga ccgcgatcga tgactgcgcg
 30 73651 acggccctcg ccccgtaagt cgactggtcg ctgcacgacg tcctggccgg
 73701 cgagggcgac cccgccctgc tggagcgggt ggacgtggtc cagcccgcgc
 73751 tgttcgccat gatggtcggg ctgtccgcgc tctggcgctc ccacggcgtc
 73801 gtcccggcgg ccgtggtcgg ccaactgcag ggcgagatcg ccgcggcctg
 73851 cgtcgccgga gccctcagcc tggccgacgc cggccgcgtg gtggcgctgc
 35 73901 gcagccaggc actgccgcaa ctgtccggac gcggcgccat gatgtcggtc
 73951 tccgcccccg tagagcgggt caccgcactc ctgccccgt ggcaggaggc
 74001 gctgtccgtc gccgcggta acggccctc gtccgtggtc gtctccggcg

74051 acaccgacgc gctcgacgcc ctgcacaccg cctgccagga acagggcgctg
74101 cgggcccgcg aggtgtccgt ggactacgcc tcgcacgggc ggcacgtcga
74151 ggccgtccgc gacgaactcg cccgcgtcct cgcgccggtc gaccgcgcgcg
74201 cccccgaggt gccgttctac tcgacggta ccggcgaccg cgtggacgac
5 74251 gccgccttcg acggcgcccta ctggtacacc aacctccgcc agaccgtccg
74301 catggaggag gccacccgcg ccttcctcgc cgccggacac cgcgtcttca
74351 tcgaggtcag cccgcacccg gtgctcgccg ccccgatcca ggagacgcag
74401 gaggccgtag cggaggccac cggcgggtcc gcggtggtcc tcggtcgcgt
74451 ccgcgcgcgac gaggcgggcc cgcggcgctt cctgacgtcg ctgcgcgagg
10 74501 cccacaccca cggcgccccg gtcgactgga ccaccacctt cgccccgtcc
74551 gcctaccagc cgggtggacct gccgacctac cccttccaac gacaggactt
74601 ctggcccagag gcccgggccg ccaccccggc cgcggcgccg gacgcgtccg
74651 acgcccgtt ctggcaactg gtcgagaacc aggacctcgc cgcgtcgcgc
74701 gacgcgtcgc gcgtccccgc cgacgacgag cacaccgcgc tcggcaccgt
15 74751 gctgccggcc ctgtccgcct ggcgcgccaa ggcccaggcc cgcacccgga
74801 tcgacgaact ccgctaccac gtccagtgga cccgggtcgc cgagcccgcg
74851 gcggccccca ccacggccg gctgctggtc gccgtcccgc cggaccacgc
74901 cgacgcccc tgggtcgccg cggcgctcga cgcctgggc accgacaccg
74951 tccgcttcga ggccaaggcc accgaccgcg cgggatgggc cgcacagatc
20 75001 gcccactcg tcgaggacgg cgaggagtcc accggcggtg tgctcgtgct
75051 ggccgcgcgc gaggatctcc acccgactt cggctcggta ccgctggggc
75101 tggggcagac cctcgtcctc gtccaggccc tcggcgacgc cggcctgacc
75151 gcgcccctgt ggtgcctgac ccgcggcgcc gtcgccaccg gccgcgacga
75201 cgccctcgac agcccgaccc agggcgccct gtggggcctc ggccgggtcg
25 75251 tggccctgga acaccccgac cgtggggcg gcctgatcga cctgcccgc
75301 accctcgacg ccgcgcgcgc ggcccgccct accggcctgc tcgccgaccc
75351 cgccggtgag gaccaactcg cgtccgcgc caccggcgtg ctgcccgcgc
75401 gcatggtgca cgccgcgcgc tccgcgcccc gaccggggcg ccgctggcgg
75451 ggccgcggca cctgcctgat caccggcggc accggcgga tcggcgggcg
30 75501 ggtcgcccgc tggatggccg agcacggcgc cgcccacctg gtcctgacca
75551 gccggcgcg cccggacgca cccggcgccg ccgcactccg ggccgaactg
75601 gaggccttg gcgcccgcgt caccctcgcc gcctgcgacg tcgccgaccg
75651 cgacgcccctg gccgccctgc tggccgacct ccccgccgac cagccgctca
75701 cctccgtctt cactccgcc ggcgaggcc acggggacgc ccgggcagcc
35 75751 gacctgaccc tagatcagct cgacgcgctg ctgcgcgcca aactgaccgc
75801 cgcccaccac ctgcacgagc tgaccgcccc cctcgacctc gacgcgttcg
75851 tgctcttctc ctccggcgcc gcggtctggg gcagcgcgcg ccagcccggc

75901 tacgccgccc ccaacgccta cctcgacgcc ctgcccggcc accgcaggtc
75951 cctcgacctg cccggcgcggt ccgtcgcttg gggcacctgg ggcgaggctg
76001 gcatggccac cgtccccgag gtccacgagc gactgcaccg ccaaggggtc
76051 cgcgccatgg aaccggacca cgcgatcggc gcgtccagc agatgtgtga
5 76101 ggacgacgac accacctctg ccgtgacctt catggactgg gaggcggttcg
76151 caccgagctt caccgcgacc cgacccagcg ccctgttcag cagggtgccc
76201 gaagccgtcc gcgcggtgac cggcgacccg ggcaccacgg ccggcgacga
76251 cgtggactcc gcgaccccg cgtcccgccg ccacctggag gagctgtccg
76301 ccgcccagcg cggccggggc ctggtcgagg cggtcgcgc cgaggcgctc
10 76351 gcgacctctg gccacgacac ccccgcgccc atccccggcg gccgtgcctt
76401 ccgcgacgtc ggcttcgact cggtcaccgc cgtcgaactg cgcaaccggc
76451 tgcgcaccgc cctcggcctg ccgtcgccgg ccgcgctcgt cttcgaccac
76501 cccacccccca cggccctcgc cggccacctc ggcgcgctgc tcttcggcac
76551 cgcccccgag gacgcgggca cggccggccc cgacgacccc gacgcccgca
15 76601 tccgcgaggc gctcgccacc gtcccatcg gacggctgcg caaggcgggc
76651 ctctcgaca tgggtgtgaa actcgccgac ggagacgga cgcgacccc
76701 cgcccccgag gccgacgccc cctcggaatc cctcgacgac atggacgccc
76751 aagccctgct gcggctggcc accgagaact cggcgaactg aaagagagct
76801 ggagcccacc atgagcacga accccgacaa gtacgtcgag gcaactccgt
20 76851 cgtccctgaa ggagatcgag cggctgcgc gccagaacga acagctgggt
76901 gcgcgggcg tcgagcccg cgcggtcgtc ggcacgggt gccgttccc
76951 cggcgggcgc acctccccg aggaacctgt ggagctggtc gccgaggggc
77001 gcgacgtcat cgggcccgtc ccgcaggacc cgcgctggga cctggagaag
77051 ttggccggcg gcggcgaggg cggcagcctc gcgcaggctc gcggcttcgt
25 77101 cgaggacgcc gccggcttcg accccggctt cttcggcac tccccgcgcg
77151 aagcggtcgc catggacctg cagcagcga tctgctgga gatcacctg
77201 gaagccctgg aacgcgcccg catcgacctg tccacctgc gcggtacccc
77251 caccggcgtc ttcgtcggca ccaccggcca ggactacggc gaggtcatca
77301 aggcgtccgc cgaggacgtc gaggtctact cgaccaccgg ccacgcggcc
30 77351 agcgtcatct ccggccggct ctctacacc ctgcccggcg agggccgggc
77401 cgtcacgctc gacaccggct gctcctcgtc cctggtcgcc ctgcaactgg
77451 ccgtccaggc gctgcgcggc ggcgagtgtc ccatggccct ggccggcggc
77501 gcgtccatca tggccacccc gggcccgttc gtcgccttca ccgcgcagag
77551 cggcctggcc gccgacggcc gctgcaagcc cttctccgac cgggcccagc
35 77601 gcaccggctg gggcgagggc gccggcatgc tggctctgat gcggctctcc
77651 gacgcccagc gcgaggggc cccggtcctc gccgtgctgc gcggctccgc
77701 catcaaccag gacggcgcc ccaacggcct gaccgctccc aacggcccct

77751 cccagcagcg cgtcatccgc gccgcgctgg acagcgccca cctcaccgcc
77801 gccgacatcg acgccgtcga ggcccacggc accggcacca ccctcggcga
77851 cccgatcgag gcccaggcgc tcctggcgac ctacggacag gaccggccgc
77901 ggcccctgtg gctcggctcg gtgaagtcca acatcggcca caccagggc
5 77951 gcctccggtg ccgccggcgt gatcaaatg atcatggcgt tgcagcgcg
78001 cgtgctgccg cgcagcctgc acgccaccga acccaccacg gacgtcgact
78051 ggaccgccgg ctccgtcgac ctctctcgac agacggtcgc ctggcccag
78101 accggacgcg ccgcgcgcgc cggcgtctcc tccttcggca tcagcggcac
78151 caacgcccac gtcattctcg aacaggcccc caccgcccc gaagagccca
10 78201 ccaccgaacc caccgtccgc ccgcgcgtcg tcccgtgggc gctctccgcc
78251 cgcaccgccg ccgccctcga cggccagcgc gccgcctca ccggccacct
78301 cgccgacacc ccgcagccgc accccctcga cgtcggctac gcgctcgccg
78351 acggacgcgc cacttcgaa caccgcgcgc tcctgctccc cgacggcacc
78401 gaactcgccc acggaaccgc cggcgaaggc ccctgcgcgc tcctcttctc
15 78451 cggccagggc tcccagcgcc cgggcatggg acgcgaactc caccgccgt
78501 tcccgtgtt cgccgcggcc ttcgacgaga tcacagcgct cctcgacacc
78551 cactcgacc gccgcgtcg cagagtcgtc tggggcaccg acgccgacct
78601 gctgaacgac accggctggg cccaaccgc cctgttcgcc gtcgaggtcg
78651 ccctctaccg cctggtcgcg tcctcggcg tcacccccga ctctcgcggc
20 78701 ggccactcca tcggcgagct cgcgcgcgc cactcgcgc gggtcctctc
78751 cctcgaagac gcctgcacc tcgtcgccgc ccgcgcccgc ctcatgcagg
78801 ccctgccgcg cggcggcgcg atgctcgca tcgcgcgcac cgaggacgag
78851 gtcaccccc acctcaccga cgacgtctcg atcgcgcgc tcaacgggccc
78901 cactccgtc gtcgtcgccg gcaccgagga agcgcgcgc gcgacgggg
25 78951 cgcgcttac cggccaggac cgcaagacca cccggctgcg ggtcagccac
79001 gccttcact cgcgcgtcat ggacccgatg ctggcggaat tccgcgcgt
79051 cgcgcgggc ctgacctacc acgagccgc catcccgtc ctctccaacc
79101 tcaccggcac cgtcgcgcgc gtcgcgcacc tgtgtccgc cgactactgg
79151 gtccgccacg tccgcgaggc ggtccgttc gccgacggc tcaccgccct
30 79201 caccgaccgc ggcgtgacca cgtcgtcga actcggccc gacggcgtgc
79251 tgtccgccat ggcccaggaa tcctgccgc acggcgccgc gcgcgtgccg
79301 ctgctgcga aggaccgcc cgaggagctc tccgcgtca ccggcctggc
79351 ccgcgcccac gtccggggc tcacggtccg ctgggcccgc ctcttcgacg
79401 gcaccggcgc gcgcgcgcg gacctgccca cctacccctt ccagcaccag
35 79451 cggttctggc cgaccgggc ccgcgcgcg caggacgtca ccgccgcggg
79501 actgggcgc gccgaccacc cgtgctcgc cgccaccgtc gaactcgccg
79551 acggggccgc ctacttgttc accagccgc tcctccgtccg gaccaccccc

79601 tggctcgccg accacggggt ccagggccg ggcctgctgc ccggcaccgc
79651 cttcgtcgaa ctggccgtcc gcgccggcga cgaggccggc tgcgaccgag
79701 tcgaggaaact gacctggcc gccccctgg tgctgcccga gcgcggcggc
79751 gtccaaactcc aggtcccggt cggcgcccc gacgcccggc gccgcccac
5 79801 cctcggcata ttctcccgcg tcgaggacgg cttcgacctg ccctggctgc
79851 aacacgccac cggcgtcctg accgcccggc ccggcgcccc cgaccccacc
79901 ttcgacgcca ccgtctggcc ccccgaggc gccgaaccg tcgacctcac
79951 cggcgcgtag gagcgcttg ccgcaactcg cttccagtag ggccccgcct
80001 tccagggcct gcgcgccgcg tggcgccgcg acaccgaggt ctacgccgaa
10 80051 gtggccctgc ccgacggcg ggacaccgac cccgcccct tcggactgca
80101 cccggccctg ctggacgccg cacaacacgc cgccgcctac gccgacctcg
80151 gcgccatcag ccgcgccggc ctgcccgttc cctgggaagg cgtctcgctc
80201 gccgcccgcg gcgccaccac cgtccgcgcc cggatcgccc cggccggcga
80251 ggacaccgtc accatcgccg tctacgacgc cgccggcggc accgtgctgt
15 80301 ccgtcgactc cctggtctcc cgcgaggtcc ccgcccgcgc acccgggccc
80351 gccggcaccg tccaccgcga ctccctcttc cacgtcgagt ggaccccgt
80401 ccagggcgcg ccgggccccg caccggccac cgtcgccgct ctcggccccg
80451 acccggaacgc cctcgccgac accctccgcg ccaccggcat ccggaccacc
80501 gccccccgcg acctggccgc cctcgccgac gccgaagggc ccgtccccga
20 80551 cctggtcgtc accacctca ccaccacccc gggcgcccc gtccccgacg
80601 ccgcgcacgc caccaccgcc gccgtcctcg ccctcgccca acagtggctc
80651 gccgacgacc gcttcgccga cgcccgccg gtccctcgta cccgcgggcg
80701 caccgacggc accgaccccg ccgcccggc cgccggcggc ctgatccgca
80751 ccgcccgcac cgagaacccc ggccgtttcg ccctcctcga cctcgcccc
25 80801 gacaccggcc ggcccgaccc cgagaccctg gccaccgccc tggccgcca
80851 ccacgacgag cccgacctcg ccgtccgcgg caccgacgtg cagcgccccc
80901 gcctggcccc tgtcccgtc gccaccgaac ccaccacctg gaaccggac
80951 ggcaccgtcc tgatcaccgg cggcaccggc ggccctggcg cggctcctgc
81001 ccgccacctg gtcgccaccc acggcgctcc ccacctgctg ctgcggacc
30 81051 gccgcccgcg gcccgccgac ggccgcccgc acctgacggc cgaactcacc
81101 gggctcgcg ccaccgtcca catcgccgcc tgcgacgtcg ccgaccggc
81151 cgccctcgcc gacctgctg gcaccgtccc ggccgggcac ccgtcaccg
81201 tcgtcgtcca caccgcccgc gtcgtcgacg acggcgctct cggctccctc
81251 accccgcagc gcctggacac cgtcctcgcg cccaaggccg acgcccctg
35 81301 gcacctgcac gaggcgaccc gccacctcga cctggacgcc ttcgtcctct
81351 tctcgtccgt cgccgccacc ctccgagcc ccggacaggc caactacgcc
81401 gccggcaacg ccttcttgga cgccctcgcc gccggcgcg ccgccaccgg

81451 cctgcccgcc acctccctcg cctggggccc gtggaccag agcgtcggca
81501 tgacaagcag cctgtccgac ctcgacgtcg agcgcacgc ccgctccggc
81551 atgccccgc tgaccctgga acagggcacc gccctcttcg acgcggccct
81601 ggccgcccgg cccgcccgc tgcgccgggt ccgcctcgac ctgcccgtcc
5 81651 tgcgcacca gggcgacatc gccccgtgc tgcgcccct gatccgcacc
81701 cccgtgcggc gcaccgccgc ccaggtctcg cagaccgccg acggcctcgc
81751 ccagcggctc gccggcctcg acgccgccgc ccgccggga gccctcctgg
81801 aactcgtccg caccagatc gcccaggtcc tcggccacgc ggacgccacc
81851 gaggtggaga ccggccgcca gttccaggac ctcggcttcg actccctcac
10 81901 cgccgtcgaa ctccgaacg ccctgaacac cgccaccggc ctgcgggtgc
81951 ccgccaccat ggtgttcgac taccgacac cacacgccct cgccgaccac
82001 ctgcgcgacg aactcctggg caccgaggcc gagtcgacca ccgcgcctcc
82051 cgtgccgacc cgtaccgccg gcaccgacga cccgatcgtc atcgtcggca
82101 tggcctgccg ctacccggc ggcatcgct caccgagga cctctggcgc
15 82151 ctggtcagcc agggcgccga cgccactggc ccgttccca ccaaccgcgg
82201 ctgggacctg gacaacctct acgacccga cccgaccgc ccgggcccga
82251 cccacgtccg cgccggcggc ttctgcacg acgccggctc cttecgagcc
82301 gacttcttcg ggatgagccc gcgcgaggcg atggccaccg actcccagca
82351 gcgcctgctg ctcgaaactct cctgggaagc cgtcgaacgc gccggcatcg
20 82401 accccgcctc actgcgcgac tccggcaccg gcgtcttcgc cggcgtcatg
82451 tacaacgact acggcaccac cctgaccggc gacgagtacg aggcgttcgg
82501 cggcaacggc agcgcgccga gcgtcgctc cgccgcgctc tcctacacc
82551 tcggcctgga agggccggcc gtcacggtgg acaccgcctg ctcttctct
82601 ctggtcgccc tgcactggc ggcgaggcg ttgcgggagg gggagtgtc
25 82651 gttggcgttg gccggtgtg tgacggtgat gtcgacgccg agcacgttcg
82701 tggagtctc gcgcgagcg ggtctggcg ctgatggcg ttcgaaggcg
82751 ttcgccgagg ccgcggacg cgtggcctgg tccgaggcg tcggcatgct
82801 ggtcctggag cggcagtcg acgcggtgc caacggtcac gagatcctgg
82851 ccgtggtgcg cggctcggcg gtcaaccagg acggtgcgtc caacggtctg
30 82901 accgcgcca acggccgct ccagcagcg gtgatccgtc aggcgttgcc
82951 cagtggcggc ctgtccacgg ccgacgtgga cgccgttgag gcgcacggca
83001 cgggtacgac gtcggtgac ccgatcgagg ccagggcgt cctggccacc
83051 tacggtcgcg accgcgaccc cgagaaccg ctgctgctcg gtcgatcaa
83101 gtccaacatc ggtcacacc aggcagcggc cgggtgcgcc ggtgtcatca
35 83151 agatggtcat ggcgatgcg cacggcgtgc tgccgcagac cctgcatgtc
83201 gacgcgccgt cctcgacgt cgattggagc gtcggcgccg tcgaactgct
83251 caccgagcag accgcctggc cggagaccg ccgggcccgt cgcgcgggtg

83301 tctcctcctt cggcatcagc ggcaccaacg cccacgtcgt catcgagcag
83351 tccccgaccg ccgtccccgc cacgcccgcg tccgccgacc ggtccgtcga
83401 ggaaccgccg gccgtcccct gggccctgtc cggcaagacc cccgacgccc
83451 tccgcgacca ggccgcccgc ctctcgtccc acgtcgaggc ccaccccgca
5 83501 ctgcgccccg tcgacatcag ctactccctg atcgccaccc gcaccgcctt
83551 cgaccaccgc gccgtcgtcc tcggcaccga ccgcgcccag gccctgcgcg
83601 ccctcaccgc cctcgccgcc ggcgagaccg acccgccgc cctcaccggc
83651 accgtccgca ccggccgcac cgccttcctc ttctccggcc agggctccca
83701 acggtccggc atgggcccgc tcctctacga gcggttcccc gccttcgccg
10 83751 aagccctcga caccgtcctc accgccctcg acgcggaact cggccacccc
83801 ctccgcgaca tcactgtggg cgaggacgct caactcgtcg accggaccgg
83851 ctacacccaa ccgcacctgt tcgccatcga ggtggcactc ttccgcctcc
83901 tggaagcctg ggcatcaca ccggacttcg tggccggcca ctccatcggc
83951 gagatcgccg ccgcacacgt cgcggcgtg ctctccctcg gcgacgctg
15 84001 ccgcctcgtc gtggcccgcg ccgtgctgat gcagtcgctg cccgaaggcg
84051 gcgcgatgat cgcgtccag gccaccgagg acgaggtcct gccctcctc
84101 accgacgacg tctcgatcgc cgcgtcaac agcccgacct ccgtcgtcgt
84151 ctccggctac gagaacgcca ccctcgcgt cgcgggcac ttccgcgacc
84201 agggccgccc caccacgcgg ctgcgcgtca gccacgcctt ccactcgcg
20 84251 ctgatggcgc cgatgctcga cgacttccgc gccgtcgtcg agagcctcac
84301 cttcaccgcc cccacgaccc ccgtcgtctc caacctgacc ggcgaactgg
84351 ccccgccga ggctctctgc tcggccgact actgggtccg gcacgtccgc
84401 gaggcggtcc gcttcgccga ccgcatccgc accctcgcg acccgggcgt
84451 caccaccttc gtcgaactcg gccccgacag cgtgctgtcc gccatggccc
25 84501 aggagtccgc cccgaaggc gccggcacca tcccgtcct gcgcccgcac
84551 cggcccgagg aacaggcgt cctggccgcc ctctgccacc tccaggtgct
84601 cggcgtcgag gccgactggt ccgccacctt ccgcggcctc gaccccgtec
84651 gcgtcgacct gccgacctac gccttcagc accgctggtt ctggcccgc
84701 gcccgaccgc ccgccccga cgacgtccgc gccgcccggc tgggcgcgc
30 84751 cgaacacccc ctctcggcg ccgccgtgca actccccgac gacgacggcg
84801 cactcttcac cggccgcctc tcctgcgca cccaccgtg gctggccgac
84851 cacaccgtcc tgggacaccg cctgctcccg ggacccgac tgggtggaact
84901 cgccgtccgc gggggcgacg agaccggcag cggccacctc gaagaactca
84951 ccctcgcgc gccctgacc ctccccgagg acggcgccac cctcctccag
35 85001 gtccgcgtcg gatccgccga cgacaccggc cgccgcaccg tcaccgtcca
85051 cgcccgcgcc gacgacaccg ccgaccgcac ctggacgctg cacgccaccg
85101 gtgtgctcgc caccacgcca ccggccgccc cggcgctcga caccacggtc

85151 tggccgcccg ccgacgccga acccctcacc accgacgact gctacgcaca
 85201 cttcaccacc caccgcttcg cctacggccc cgccttccag ggcctgcggg
 85251 ccgcctggcg cgcggcgac gtgctgtacg ccgaggtcgc cctgccggag
 85301 tccgccaccg acgaagcggc cgccttcggc ctgcacccgg cgctcctgga
 5 85351 cgccggcctg cagccgcgcg tcctcgccga cgaccgcgac accggactcc
 85401 cgttctcctg ggaaggcgtc actctgcacg cctccggcgc caccgcgcta
 85451 cgcgtccggc tcgccccgaa cggccccaac ggcctgtccg tcaccgccgc
 85501 cgaccggcc ggcaaccccg tcgccaccgt caccgcctg ctgcccgc
 85551 ccctggacgc cgagcagttg accatccaca gcgccctgac ccgcgacgcg
 10 85601 ctcttcacc tggactggac cccggtccc cttcccgaca ccgccaactc
 85651 cgcgcgcgcg gccctcctcg gcccgacac cgccgtgctc gccgacgccc
 85701 tcggcgaccc ggccgtcgca cgccacgcaa ccctcgacga cctcctggcc
 85751 ggggacacca cccgcccgc caccgtctc gtccccctcg gcgcccact
 85801 cgacggcgac accgcgcagc acgcgcacgc cctcaccgc agcgcgctga
 15 85851 ccctcgcca gcagtggctc gccaccgacc gcctcgccga ctccgcctg
 85901 gtcttcgtca cccacggagc cgtcgccacc gacgacgcgc cccccaccga
 85951 cctggccgcg gccgcggtct ggggcctgat ccgctccgcg cagaccgaga
 86001 accccggcac cttaccctc ctcgacctcg acaccgagcc cgactcgacc
 86051 accgcgctca gccgcgccct gaccctcgac gaaccacagc tcctcctccg
 20 86101 cgccggccgc gccgcgcgcg cccgcctcac ccgaccccc gcccccacca
 86151 ccaccacca caccgcgtgg tccgcggacg gaacgggtgt ggtgacgggt
 86201 ggtacgggtg gtctgggtgg gttggtggcc cggcatctgg tgcggtcgtg
 86251 tggggtgcgg catttgttgt tgaccagtcg ttctggtgtg ggtgctgcgg
 86301 gtgcggcccg gttggtcgcg gagttggagt cgttggcgcc gcgggttgtg
 25 86351 gttgcggcgt gtgatgtggg tgatggctcg gctgttcgg agttggtgc
 86401 cgggtgtgct gagtcgtatc cgttgtctgc ggtggtgcat gcggctggtg
 86451 tgttgatga cgggtgtgtg ggttcgttga cgcggagcg gttggtgcg
 86501 gtgttcgctc cgaagggtga tgggtgcgtg aacctgcatg aggcgacgcg
 86551 tggcttgat ctggacgcgt ttgttgtct ctcgtctgtt gcgggtgtgt
 30 86601 tcgggggtgc gggtcaggcc aactatgcg cgggtaatgc gtttttgac
 86651 gcgttgatgg ttcacgggt ggctggtgg ttgcctggtg tgcgttggc
 86701 gtggggtgct tgggatcagg gtgtgggat gacggcggg ctgacggagc
 86751 gggatgtccg tcgtgctgct gagtcgggta tgccgttgtt gacggttgat
 86801 cagggtgtgg cgttgttca tgccggttg gcgacgggga gtgccgcgtt
 35 86851 ggtgccggtc cgtctggacc tggccgact gcgcaccgg ggcgacatcg
 86901 caccgctcct ccgcggcctc gtccgcgcac cgctgcgcc caccgcggcc
 86951 accggcctcg ccaccggcgc ggacaccggc ctcgtccaac ggctcggccc

85151 tggcgcgccg ccgacgccga acccctcacc accgacgact gctacgcaca
85201 cttcaccacc caccgcttcg cctacggccc cgccttccag ggcctgcggg
85251 ccgcctggcg cgcggcgac gtgctgtacg ccgaggtcgc cctgccggag
85301 tccgccaccg acgaagcggc cgccttcggc ctgcaccggc cgctcctgga
5 85351 cgccggcctg cagccgcgc tcctcgcga cgaccgcgac accggactcc
85401 cgttctcctg ggaaggcgtc actctgcacg cctccggcgc caccgcgcta
85451 cgcgtccggc tcgccccgaa cggccccaac ggcctgtccg tcaccgccgc
85501 cgaccgggcc ggcaaccccg tcgccaccgt caccgcctg ctgcgccgcc
85551 ccctggacgc cgagcagttg accatccaca gcgccctgac ccgcgacgcg
10 85601 ctcttcacc tggactggac ccgggtccc cttcccgaca ccgccaactc
85651 cgcgcgcgc gccctcctcg gcccggaac cgcctgtcgc gccgacgcc
85701 tcggcgaccc ggccgtcgca cgccacgcaa ccctcgacga cctcctggcc
85751 ggggacacca cccgcgccgc caggtcctc gtccccctcg gcgccccact
85801 cgacggcgac accgcgcagc acgcgcacgc cctcaccgc agcgcgtga
15 85851 ccctcgtcca gcagtggctc gccaccgacc gcctcgcga ctccgcctg
85901 gtcttcgtca cccacggagc cgtcgccacc gacgacgcgc cccccaccga
85951 cctggccgcc gccgcggtct ggggcctgat ccgctccgcg cagaccgaga
86001 accccggcac cttaccctc ctgcacctg acaccgagcc cgactcgacc
86051 accgcgtca gccgcgccct gaccctcgac gaaccacagc tcctcctccg
20 86101 cgccggccgc gccgcgccg ccgcctcac ccgcacccc gccccacca
86151 ccaccacca cagccgtgg tcgcggacg gaacggtgtt ggtgacgggt
86201 ggtacgggtg gtctgggtgg gttggtggc cggcatctgg tgcggtcgtg
86251 tggggtgcgg catttggtgt tgaccagtcg ttctggtgtg ggtgctgcgg
86301 gtgcggccgg gttggtcgc gagttggagt cgttgggcgc gcgggttgtg
25 86351 gttgcggcgt gtgatgtgg tgatggctcg gctgttcgc agttggtgc
86401 cgggtgtcgc gagtcgtatc cgttgtctgc ggtggtgcat gcggctgggt
86451 tgttgatga cgggtgtgtg ggttcgttga cgcggagcg gttggtcgc
86501 gtgttcgctc cgaagggtga tgggtcgtgg aacctgcatg aggcgacgcg
86551 tgggtctgat ctggacgcgt ttgttgtctt ctgctctgtt gcgggtgtgt
30 86601 tcgggggtgc gggtcaggcc aactatgcg cgggtaatgc gtttttggac
86651 gcgttgatgg ttcacgggt ggctggtggg ttgcctggtg tgcgttggc
86701 gtggggtgct tgggatcagg gtgtgggat gacggcggg ctgacggagc
86751 gggatgtccg tcgtgctgct gagtcgggta tgccgttgtt gacggttgat
86801 cagggtgtgg cgttggtcga tgcggcgtt gcgacgggga gtgccgcgtt
35 86851 ggtgccggtc cgtctggacc tggccgact gcgcaccgg ggcgacatcg
86901 caccgtcct ccgcggcctc gtccgcgcac cgctgcgcg caccgcggcc
86951 accggcctcg ccaccggcgc ggacaccggc ctgctccaac ggctcggccg

87001 actcgaccac gcccaacgcc acgaggcact gctcgacatg gtccgcagca
87051 gcgcgcgct cgtcctcggc cagcccgacg gcaacgccat cgacgccgaa
87101 cgcgccttcc gcgacctcgg ctctcgactcg ctcaccgcgg tcgaactccg
87151 caaccgtctg cgcaccgccca ccggcctgca cctgtcggcc accatggtct
5 87201 tcgaccaccc caccctgtcc gccctcggcg agcacctgcg ggacgagttg
87251 ttcggcgcgg tcgagagcga ggtgcgggtg ccggtccagg cactgccgcc
87301 gaccgccgac gatcccatcg tgggtggtgg catggcctgc cgtttccccg
87351 gtggtgtgac ctgccccgag gacctgtggc gcctggtcga cgacggcacc
87401 gacgccatca ccaccttccc gaccaaccgc ggctgggacc tggacaacct
10 87451 ctacgacccg gaccccgagc acttcggcac gtcgtacacc cgctccggtg
87501 gcttcttgca cgaggcgggg gagttcgacc cggcgttctt cggaatgagc
87551 ccgcgtgagg cgctggcaac cgactcccaa cagcgtctcc tgetggaatc
87601 ctcttgaggag gcgatcgagc gggccggcat cgacccgctg accctgcgcg
87651 gcagcggcac cggcgtcttc gccggcgtga tgtacagcga ctacgggagc
15 87701 atcctcggcg gcaaggagtt cgagggttc caaggccagg gaagtgcggg
87751 cagcgtggcc tcgggccgcg tctcctacgc cctcggttc gagggcccgg
87801 ccgtcacggt ggacacggct tgetcttctt ctctggtcgc cctgcactgg
87851 gcggcgagg cgttgccggc gggggagtgc tcgttgcggt tggccggtg
87901 tgtgacggtg atgtcgacgc cgagcacgtt cgtggagttc tcgcggcagc
20 87951 ggggtctggc gcctgatggt cgttccaagg cgttcgccga ggccgcggac
88001 ggcgtcggct ggtccgaggc cgtcggcatc ctctcctgg agcggcagtc
88051 ggacgcgggt cgcaacggcc acgagatcct cgccgtgatc cgcggctcgg
88101 cggtaacca ggacggtgcg tccaacggcc tgaccgcgcc caacggcccg
88151 tcccagcagc gcgtcatccg tcaggcgttg gccagtggcg gcctgtccac
25 88201 ggccgacgtg gacgccgtcg aggcgcacgg cacgggtacg acgctcgggtg
88251 acccgatcga ggcccaggcg ctcttgcca cctacggccg tgaccgcgac
88301 cccgagaacc ccctgtggtt gggctccctg aagtccaaca tcgggcacac
88351 ccaggcagcg gccggtgtcg ccggtgtcat caagatggtc atggcgatgc
88401 ggcacggcgt gctgccgag accctgcatg tcgacgcgcc gtcctcgac
30 88451 gtcgattgga gcgtcggcgc cgtcgaactg ctcaccgagc agaccgcctg
88501 gccggagacc ggccgggtcc gtcgcgccgg tgtctcctcc ttcggcatca
88551 gcggcaccaa cggccacgtc atcgtggaac agccggcgct cgtcgaaagc
88601 ccggccgcgg agccgagcgg acgcgaacct ggcgtcgttc cgtgccgct
88651 gtccggaaaag tccccgagg ccctgcgcga ccaggccgca cgctgctgg
35 88701 ccgggttggc ggagcggccc gcgtgcgcc cgtcgcacct cggctactcg
88751 ctggcgacga cccgttcggc gttcgaccac cgggcgggtg tgctcgccac
88801 cgaccgcgcc gatgcggtcc gcgcgctgac ggcgtcgc ccgccgcagc

88851 cggatctctc cgccgtcgtc ggcgacaccc gcacgggtcg tcacgcggtg
88901 ctgttctcgg gtcagggctc gcaacgcctg ggcatggggc gtgagttgta
88951 cgagcgtttc ccggtcttcg ccgaggctct cgatgtcgcg atcgaccacc
89001 tggacgccgc cttgcccgc caggccagtc tgcgtgaggt gatgtggggc
5 89051 gacgatgtcg agctgctgga cgagacgggt tggacgcagc cggctctgtt
89101 cgccgtcgag gtggccctgt tccggtcgtt ggagagttgg ggtgtccgtc
89151 cggacttcgt ggccggtcat tccatcgggt agatcgccgc ggcgcagtgc
89201 gtcgggggtgt tctcgtgga ggacgcctgc cgtctggtgg ccgcccgtgc
89251 gacgctgatg caggcgtcgc cgaccggcgg cgcgatgacg gcgatccagg
10 89301 ccgccgagga cgaagtcacc cagcacctga ctgacgacgt ctcgatcgcc
89351 gccgtcaacg gcccgacctc cgtggtcgtc tccggggccg agagcgtgc
89401 ccgcacggtg gccgaccggc tcgcggagaa cggccgcaag acgaccgggc
89451 tgcgggtctc gcatgcgttc cactcgccgt tgatggatcc gatgctggcg
89501 gagttccgtg cgggtggccga gggcctgtcc tacgccaccc cgacctccc
15 89551 cgtcgtctcg aacctgacgg gccggtcggc caccggccgat gacctctgct
89601 cggccgagta ctgggcgcgc cagtcgcgcg aggcgggtccg cttcgccgac
89651 ggcgtcagca ccctggagaa cgagggcgtc accacgttcc tggaactggg
89701 accggacggc gtgctgtccg ccatggccca gcagtcgctc accggcgacg
89751 ccgccaccgt ccggccctc cgcaaggacc gcgacgagga gacgtccgcg
20 89801 ctcaccgccc tcgcccacct ccacacggca ggtctccgcg tcgactgggc
89851 ggcgttcttc gccggcagcg gcgccacccg cgtggacctg ccgacctacg
89901 ccttcagca cgccacctac tggccacccg gcacctgcc caccgcgcac
89951 gccgcgccg tcggcctcac cgccgccgag caccgctgc tgaacggtc
90001 ggtcgaactc gccgaaggcg aaggggtgtt gttcaccgga cggctgtcac
25 90051 tgcagtcaca tccgtggctg gccgaccacg ccgtcatggg acaggctctg
90101 ctgcccggca ccgcaactgt ggaactggcg ttccgggccg gcgacgaggc
90151 cggttgcgac cgcgtcgagg aactgacgt cgcgcacccg ctcgtcctgc
90201 ccgagcgcgg tgcggtacag acccaggctc gggtcggcgt cgccgacgac
90251 accggccgcc gtaccgtcac cgtccactcc cggcccagc acgcgaccga
30 90301 cgtgtcgtgg acccagcacg cgaccggcac cctgaccatg ggctccgccc
90351 cggccgacac cggtttcgac gccactgcct ggccgcccgc cgacgccgaa
90401 cccctcgcca ccgacgactg ctacgcgcgc ttacgacgc tcggcttcgc
90451 ctacggggcg gtcttcagg gcctgcgggc cgcctggcgc gccggtgacg
90501 tgctgtacgc cgagggtggc ctggcgaggt ccaccggcga cgaggcgacc
35 90551 gccttcggtc tgcaaccgc actgctcgac gccgccctgc acgcctccct
90601 cgtcgccac gagggcgagg agagcaacgg cggactgccg ttctcctggg
90651 agggcgcgac cctctacgcg accggcgcca ccgcgctgc cgtccggctg

90701 accccgacgg gcaccgacgg ccgttcgggtg gccatcgccg tggccgacac
 90751 cgccgggtcgt ccgggtcgccg ccatcgacaa cctcgtctcg cgccgggtct
 90801 ccggcgacca gttgaccggc gccgcgggac tggcccgca cgccctgttc
 90851 accctggact ggaaccccg accggagaac ctcgtaccgg agaaccctgt
 5 90901 accggagaac accggcgggg gccacgcca ggaccaggac ggccggccccg
 90951 ccgcggccac cgtcgcgctg gtcggcgcg accgcaccgc gatcgccgcc
 91001 gaactgaccg ccgccggcat ccacaccacc ctccaccccg acctcaccac
 91051 cctcgccacg accgacgccg acgttcgaa gacggctctc atccccctca
 91101 ccggaaccgg aaccggaacc ggacccggga ctgagtcgac ggacggaatc
 10 91151 gggacggggg ccgccgagtc ggacgcgtcc gccccctccc cggccgaggt
 91201 cgcccacacc ctgtccaccg ccgcactcgc cctcgtccag gagtggaccg
 91251 cacaggagcg cttcgccggc tcccgccctgg cgttcgtcac gaccggggcg
 91301 acggccgccg gcggtaccga cgtcatggac gtggccgccg ccgcggtctg
 91351 gggcctggtc cgatccgccc agtccgaagc cccggacacc ttcgtcctga
 15 91401 tcgaccgtga ccccgcccg gccggcacgc acgaccgcac agccgcccgc
 91451 gaacggggcc aactgctcct acgggcaactg cacaccgacg aaccgcagct
 91501 cgccctgcgt gacggcgggc tgctcgccgc ccgcctggcc cgcttcgaca
 91551 ccgcgccgc gctcaccgc ccggccgacc gggcctggcg gctcgacagc
 91601 acggccaagg gcagcctcaa cggcctcgcc ctgaccccg atccggcggc
 20 91651 actggcgccg ctcaccggcc acgaggtgcg ggtcgaggtg cgtgccgcg
 91701 gcctgaactt ccgtgacgtg ctcaacgcgt tgggatgta tccgggtgat
 91751 gatgtcggat cgttcggttc ggaggcgcc ggtgtggtcg tcgaggtcgg
 91801 accggaggtg accggcctgg ccccgccga ccaggatcat ggcatgatca
 91851 ccggcagctt cggctcgctc gccgtggacg acgcgcggcg cctcgccgc
 25 91901 ctgcccagg actggtcctg ggagacgggt gcgtcggtgc cgttggtgtt
 91951 cctcaccgcg tactacgccc tgaaggagtt ggggtggtctg cggcgggggg
 92001 agaaggtgct ggtgcatgcc ggtgccggtg gtgtcggtat ggcggcgatc
 92051 .cagatcgccc ggcatgtcgg tgccgaggtg ttcgccacgg ccagtggagg
 92101 caagtgggac gtgttgcgct cgctcgcggt ggccgacgac cacatcgctt
 30 92151 cctcccgcac cctcgacttc gaggcggcct tcgccgaagt cgccggcgac
 92201 cgcggcctgg acgtcgtcct caactccctc gccggtgact tcgtcgacgc
 92251 ctcgatgcgg ttgctcgcg accggcgccg gttcctggag atgggcaaga
 92301 ccgacatccg cgcccgacg tccgttccc accgcctctc ctaccagtcc
 92351 ttcgacctcg cctgggtggt gccggaaacc atcggcacca tgctggccga
 35 92401 gctgatggac ctcttcgca ccggcgcaact gcggccactc ccggtccgca
 92451 cctgggacgt ccggcacgcc aaggacgcgt tccgcttcat gagcatggcc
 92501 aagcacatcg gcaagatcgt gctcaccctg ccccgctcct ggaagcccga

92551 gggaacggtg ttggtgacgg gtggtacggg tggctcgggt gggttggtgg
92601 cccggcatct ggtgcggtcg tgtggggtgc ggcatttgtt gttgaccagt
92651 cgttctggtg tgggtgctgc ggggtcggcc gggttggtcg cggagttgga
92701 gtcgttgggc gcgcgggttg tggttgcggc gtgtgatgtg ggtgatggct
5 92751 cggctgttgc ggagttggtt gccggtgtgt cggagtcgta tccgttgtct
92801 gcggtggtgc atgcggctgg tgtgttggat gacggtgtgg tgggttcgtt
92851 gacgccggag cggttggctg cgggtgttgcg tccgaagggtg gatggtgcgt
92901 ggaacctgca tgaggcgacg cgtggtctgg atctggacgc gtttgttgc
92951 ttctcgtctg ttgcgggtgt gttcgggggt gcgggtcagg ccaactatgc
10 93001 ggcgggtaat gcgttttttg acgcgttgat ggttcacggtg gtggtcgggtg
93051 ggttgccctg tgtgtcgttg gcgtggggtg cttgggatca ggggtggtgg
93101 atgacggcgg ggctgacgga gcgggatgtc cgtcgtgctg ctgagtcggg
93151 tatgccgttg ttgacgggtg atcagggtgt ggcgttggtc gatgcggcgt
93201 tggcgacggg gagtgcccg cgtgtgccgg tccgtctgga cctggccgca
15 93251 ctgcgcaccc ggggcgacat cgcaccgctc ctccgcggcc tggtaaggc
93301 gcccattccg cgcgagccg ccaccacacc cggcgacacc ggactcggc
93351 agcagctcac ccggtccag cgcgccgagc gacgggacac cctcctcgcg
93401 ctgctcccg accaggccgc gatggtcctc ggccacacct cgggcgacgg
93451 cgtcgacctg tcccgcgct tccgcgacct cggcttcgac tcgctcaccg
20 93501 cggtcgaact ccgcaaccgc atcggcgcg ccaaccggcct gcggctaccg
93551 gccacggccg tcttcgacta ccccaccgcc gatgcctcgc ccgcacacct
93601 gctcaccgaa ctgctcgccc ccgacgccga gtcggacccc gacgagcccg
93651 gcgacccac cgcgggaccg accgacgacc ccatcgtcac catcggcag
93701 agctgcccgt tcccggcgga catcggtcgc ccggaggacc tgtggcgcc
25 93751 gctcggcgac ggcgccgacg tcgtcaccga cttcccgacc aaccgcggct
93801 gggacctgga caacctctac gaccccgacc ccgcgcacgc cggcacctcg
93851 tacgcccga ccggcggttt cctgcacgac gccgccgact tcgacgccga
93901 cttcttcggc atgagcccc gcgaggccat ggccacggac tcccagcagc
93951 gcctgctgtt ggagtctcgc tgggaggcga tcgagcgggc cggcatcgac
30 94001 ccgctgacct tgcgcgacag ccgcaccggc gtcttcgccc gcgtcatgta
94051 cagcggctac ggcaaccgcc tcgacggcgc cgaattcgaa ggcttcagg
94101 ggcagggcag cgactgagc gtggcctccg gccgggtctc ctacaccttc
94151 ggcttcgaag gcccggccat gacggtcgac accgcctgct cctcctcgct
94201 ggtcgccctg cacctcgccc cacaggcact ccgcggcggt gagtgcacc
35 94251 tcgccctcgc cgggtgtgtc accgtgatgt ccatcccga caccttcac
94301 gaggttctccc ggcagcgcg actggcccc gacggccgct ccaagccgtt
94351 ctccgagtc gccgacggcg tcggctggtc cgagggcgtc ggaatgctgc

94401 tcctggagcg ccagtcggac gccgtgcgca acggccacca gatcctggcc
94451 gtggtgcgcg gctcggcggg caaccaggac ggtgcgcca acggcctgac
94501 cgcgccaac ggcccgctcc agcagcgggt gatccgtcag gcgttgcca
94551 gcggcggcct gtccacggcc gacgtggacg ccgtcgaggc gcacggcacg
5 94601 ggcaccacgc tcggtgacct gatcgaggcc caggccctcc tggccacct
94651 cggccgacgc cgcgaccccg agaaccgct gctgctcggg tcgatcaagt
94701 ccaacctcgg ccacacccag gcagccgccg gtgtcgccgg cgtcatcaag
94751 atggtcatgg cgatcgcgca cggcgtgctg ccccgagcc tgaacatcac
94801 cgagccgtcc tcgcacgtcg attggagcgc cggcgccgtc gaactgctca
10 94851 ccgagcagac cgcttgccg gagaccggcc gggcccgctc cgccggtatc
94901 tcctccttcg gcatcagcgg caccaacgcc cactcatcc tggagcagcc
94951 ggaggccgcg cggcactcgg cgccggaaga agccgacacg gcggaggcag
95001 ccgccaaggc gccggccacc gcgcacctgc ccgtaatgcc gtgggcactg
95051 tccggcaaga cgccggaggc cctgcgtgcc caggccgcac gcctcctcgc
15 95101 ccacctccag cagcgcccg aactcgcacc cgccgacatc gccctgtccc
95151 tcgccacca gcgctcccag ttcaccacc gggcagtcgt cctgagcacc
95201 gaccgtgacg aggcgacccg cgcgctgtcc gccctcgcca ccaccgccgc
95251 gtccgacccc tcggccctca ccggcacggg caccatggga cgttgcgcg
95301 tgctgttctc gggtcagggc tcgcaacgtc tgggcatggg gcgtgagttg
20 95351 tacgagcgtt tcccgtctt cgcgagggt ctcgatgtcg tgatgatca
95401 cctggacgcc gccttgccc cccaggccgg tttgcgtgag gtgatgtggg
95451 gcgacgatgt cgagttgctg aacgagacgg gttggacca gcccgcgctc
95501 ttcgccatcg aggtggcgt gtttcggctg gtggagagtt ggggtgtccg
95551 tccggacttc gtggccggtc attccatcgg tgagatcgcg gcggcgcatg
25 95601 tcgtcggggg gttctcgttg gaggacgct gccgtctggt ggcgcgcgg
95651 gcgacgctga tgcaggcgtt gccggccggg ggcgcgatga tcgcggtcca
95701 ggcgaccgag gacgaagtca tccgcacct gaccgacgag gtggcgatcg
95751 cggccgtcaa cgcccgaac tccgtggtga tctcgggcgc agaagaggcc
95801 acgcagaccg tggcacaaca cttcgccgac caggggcgcc ggacgaccgc
30 95851 gctgcgggtc tcgatgcgt tccactcgcc gctgatgatg ctggcgaggt
95901 tccgtgcggg ggcgagggc ctgtcctacg ccaccccgac cctccccgtc
95951 gtctcgaatc tgacgggcca ggtggccacg gccgacgaac tctgctcggc
96001 cgagtactgg gtgcgccacg tccgtgaggc ggtccgcttc gccgacgggtg
96051 tgacggccct cgaagccgag ggcgtgcgga ccttcctgga actcggcccg
35 96101 gacggcgtcc tcgcccgcgt gccagggaa accgtcgccg acgacacggt
96151 caccgtcccc gtctccgca ggaacatgcc cgaggaacgg accctgctca
96201 ccgcaactcg ccggctccac accaccggaa ccccgatcga ctgggcccgc

96251 ctccctggccc cgaccggcgc ccgcccgggtg gacctgccga catacgcgtt
96301 ccaacaccgt cccttctggc cctccggccc ccgcgacacc gcggatgccg
96351 ccgccgtcgg catcgccggc gcgagccacc cgctcctcaa cggcatcgtc
96401 gaactcgccg acgaagaggg cctgttggtt accggacggc tgtcactgca
5 96451 gtcgcatccg tggctggccg accacgccgt catgggacag gtctgctgc
96501 ccggcaccgc actgctgga ctcgccctgc gcgccggcga cgaggtcggc
96551 tgtgaccacg tcgaggaact gacgctcgcc gcaccgctcg tcctgcccga
96601 gcgcggcgcg gtacagaccc aggtccgggt cggcgtcgcc gacaccaccg
96651 ggccgcgcac cgtcacgate cactcgctc ccgcacgcgc caccgacc
10 96701 gacagtgaca cccacaccgg caccgacacc ccgtggaccc aacacgccac
96751 cggcgtcctc gtcgccggcc tgcggcgac ggcaaccgtc ccgttcgatg
96801 ccaccgtgtg gccgccggcg caccgcaac ccgttgacct ggccgacttc
96851 tacgcgtccc gggccggcga aggattcggc tacgggccc ctttccaggg
96901 cctgcgagcg gcctggcgcc gcgacggcga ggtgttcgcc gatgtcgac
15 96951 tgccggaggc cggccgtacc gaagccgagg cgtacgggct gcatccggca
97001 ctgctcgacg ccggactgca cgcagcctgg ctgctcgccc cggacgggga
97051 gcccacacgg acgggcagcg tgccgttctc gtggcgcgcc gtcttctcgg
97101 ccgttcccg tgccctctcg gtccgcgtcc gactcggccg cgactccgac
97151 ggaacgctga gcctggccat cggcgacacc accggtgcac cggtcgcgtc
20 97201 cgtacaggcc ctctccatgc gcaccgtctc ggtgacggcc ctgagcgcca
97251 ccgcgggcct cggccgcgac gcgctgttcc gcctggactg ggccctcgcc
97301 ccggagccgg cgtgccagcc ggacgacacg gtgaccgtga tcccggcggt
97351 cgcggtcgtc ggtacgaaa cctccgaact cacctccgag ctcaccgcgg
97401 ccctgcgtgc ggcggcgcc gacgtcgacg tccgcacgac cctgtcgacc
25 97451 gatgaaccgg cgccgcgct gatcgcgctg ccgctcgtcg cctccgacca
97501 gaccggcacc gcagaagcgg caccgctccc ggccggcgtg caccacctca
97551 cccgcgagc cctggccctc gtacagaccc gcctgcaaga gcagcacttc
97601 gcggacacga agttcgtctt cgtgaccggt ggtgcgacgg tcgggcgtga
97651 tgtggtgct gctgcgggtg ggggtctggt gcgttcggcg cagtcggaga
30 97701 atccgggttg ttttgcctcg gtcgatctgg atccggatgg tgcggtggt
97751 gcggctgcgc tcgtcgctgc gttggtcagt ggtgagccgc agcttgcggt
97801 gcgcggtgat gtgttcggg tcgcgcgtct ggtgcggcgg ccgctcaccg
97851 aggtcgggtgc ggggtctgat ggcaccgggg atggcgctcg gggtggtct
97901 ggtgtgctgt tctcgggtga ggggtcggtc ctggtcactg gtggtacggg
35 97951 tggctcgggt gcggtgttg cgcgtcatct ggtggccgag tatggggtgc
98001 gggatctgct gttggtcagt cgcagtgtg aacgtgccgt ggggtgctgg
98051 gagttggtgg cggagcttgc ggggtgtgggt gcgcgggtgc ggggtggtgc

98101 gtgtgatgtg accgatcgtg ccgcggtggt ggagttggtt ggcgggcatg
98151 cgggtgtccgc ggtggttcat gcggctggtg tgctggatga .cggcatggtg
98201 ggtgcgttga ccggggagcg gttgtccgcg gtgctgcggc cgaagggtgga
98251 tgctgtcttg catctgcatg aggcgacccg cggcctggat ctggacgcgt
5 98301 tcgtcgtctt ctcctctctc gccggggtct tcggcagtcc cggccaggcc
98351 aactacgcgg ccgcgaacgc cttcctggac gcgctgatga cgcggcgccg
98401 ggcggaggga ctgcccggcc tgtcactcgc atggggaccg tggtcgctga
98451 ccgatggcac atcgggcatg ctgcggacg ccgaggccga tcgcctgacc
98501 cgttcgggag tgccaccgt gaccgcggag caaggactgg cactgttcga
10 98551 cgcggccctg gcgaccgtg acgccacctg cgtcccggtc cgcctggacc
98601 tgcggcact gcgtgccag ggtgaggtgc cgcccttgct gcggctccctg
98651 atccgaggcc gtcgcgcgcg cgcgcgcgcg gcggaatcgg caaccgccac
98701 cgggctgcgg gaacggctcg tcggactgaa cccggtcgag cgacaggaag
98751 tcctcctgga cctcgtgcgc ggccagggtg ccctggctct cggccacgcc
15 98801 gacgcgcacg acgttcaccc ggcacgtgcc ttcagggagt tgggcttcga
98851 ctccctcacc tcggtggaac tgcgcaaccg cctcaacacc gtcaccggtc
98901 tgcggctccc cgcgaccatg gtgttcgact atccgaccgt cgaagtgtg
98951 gtctcctacg tcctggacga gttgtgggc acggatgccg aggtggcgac
99001 cgtgcagccg gccgcggtg cggtggcgga cgatccgatc gtcacgtg
20 99051 gcatggcctg ccgtacccc ggcggggtcg cctctccgga cgacctgtg
99101 aggtcgtca cggacggcgt cgacgcggtg tcccccttc cgaccaaccg
99151 tggttgggac gtcgaatccc tctatcccc ggaccccgac catctcgta
99201 cctcctacac gcgctcgggt gggttcctgc atgaggcggg ggagttcgat
99251 ccggggttct tcgggatgag tccgcgggag gcgttggcga ccgattccca
25 99301 gcagcggttg ttgttgagt cgtcgtggga ggcgatcgag cgggccggta
99351 ttgatccggt gaggttgcg ggtagtcgga cgggtgtgtt cgcgggggtg
99401 atgtacagcg attacagcgc gatgttggcg agtccggagt tcgagggtt
99451 ccagggcagt gggagttcgc cgagtttggc gtcgggtcgg gtggcctaca
99501 cgttgggggtt ggaaggcccg gcggtgacgg tggatacggc gtgttcgtcg
30 99551 tcgttggtg cgatgcactg ggcgatgcag gcgttgcgta gtggtgagt
99601 tgggttgcg ttggccggtg gtgtgacggt gatgtcgacg cctgcgggtg
99651 ttgtggactt tgctcggcag cggggtttgt cgcgggatgg tcggtgcaag
99701 gcgtttgcgg atgcggccga tgggtgtggc tgggtccgagg gcgtcggcgt
99751 gttggtcctg gagcggcagt cggacgcggt gcgcaatggt cacgagattt
35 99801 tggctgtggt gcggggttcg gcggtcaacc aggatggtgc gtccaatggt
99851 ttgacggcgc ccaatggtcc gtcgcagcag cgggtgatcc ggcaggcgtt
99901 ggccagtggg ggtctgacgg ccggtgacgt ggatgtggtg gaggcgcatg

99951 gtacgggtac gacgctcggg gatccgatcg aggcgcagggc gttgttggcg
100001 acgtatgggc gggatcgtga gcctgagcgg ccgttggtgt tgggttcggg
100051 gaagtcgaat ctggggcata cgcaggctgc tgcgggtgtg gcgggcggtta
100101 tcaagatggg gttggcgatg cggcatgggtg tggtgccgcg gacgttgcat
5 100151 gtggatgcgc cttcttcgca tgtggactgg tccgaggggtg cgggtggagct
100201 gctcagttag caggcggcct ggccggagac gggtcgggtg cggcgggcgg
100251 gtgtctcctc cttcggcatc agcggcacca acgcgcacgt gattctggag
100301 cgtccggagg ccgcgcggcg tccgggtgat gagacgaaca ccgtggagcc
100351 gtccacgggtg ccgtgggtgc tttccggcaa gacaccggag gccctgcgcg
10 100401 cccaggcggc gaaactgttg tcgtcgatcg aggaacgccc ggagcttcgc
100451 ctgctcgacg tgggaatgtc cctggtcacc ggccgctcga ctttgaaca
100501 ccgtgcgggtg gtactcgccg ccgaccgtgc cgacccgcc cgcgattgt
100551 cggcgattgc tgccgacgag gccgatgctg ccgctgccac cggccgcgtc
100601 ggcgcgggtc gtcacgcggg gctgttctcg ggtcaggggtg ctcaactct
15 100651 gggcatgggg cgtgagttgt acgagcgttt cccggctctc gccgaggctc
100701 tcgatgtcgt ggtcgaccac ctggacgccg ccttgcccgc ccaggccggg
100751 ttgctgagg tgatgtggg cgacgatgcg gagttgctga acgagacggg
100801 ttggacgcag ccggctctgt tcgccatcga ggtggcgctg ttccggctgg
100851 tggagagttg ggggtgtcgt ccggacttcg tggccgggtca ttccatcggt
20 100901 gagatcgcg cggcgcagt cgccgggggtg ttctcgctgg aggacgcctg
100951 ccgcctgggt gccgcccggt cgacgtgat gcaggcggtt cggccggcg
101001 gtgcgatgat cgcgggtccag gcgaccgagg acgaagtcac cccgcactg
101051 accgacgacg tagcgatcgc cgccatcaac gggccgaacg cactggtcgt
101101 atcgggtgtg gaggatgccg ccgtcgagat cggggcgcgg ttccgccc
25 101151 aggggcgtcg caccacccga ctccatgtgt cgcgtcggtt cactcgccg
101201 ttgatggatc cgatgctggc ggagttccgt gtggtggcg agggcctgtc
101251 ctatgctgct ccgtccctcc ccgtcgtctc gaatctgacg ggccagggtg
101301 ccacggccga cgaactgtgc tcggccgagt actgggtgcg ccacgtccgc
101351 gaggcgggtc gtttcgccga cggggtgacg gccctcgaag ccgagggcgt
30 101401 gcgacacttc ctggaactcg gcccgacgg cgtcctcgcc gccatggcag
101451 gagcctcgt caccgaatcc tccctcgcg taccgtgct ccgtaaggac
101501 cggccggagg aaccggcggc actcgccgcc ctggcccagt tgcacatgc
101551 cggcgcgcg gtcgattggc ccgtgctctt cgctgggtgt ggtgcggggc
101601 ggggtgagtt gccgacgtat cgttccagc gtgggtggt ctggccggtt
35 101651 ggtcgggttg gtgttggtg tgatgtgggt gctgtggggc ttgggtctgc
101701 ggggcatccg ttgttggtg ctgcgggtga gttggctgc ggtgcggggg
101751 tgggtgtgac gggtcgtctg tcgtgtcgt cgcatggtt gttggctgat

101801 catgcggtga tggggcggggt gtttgttcct ggtacggcgt tgctggagat
 101851 ggtgatgcgt gctggtgatg aggtgggggtg tggctcgtgtt gaggagctga
 101901 cgttggcggc gccgttggtg ttgcctgagc gtgggtgggt gcgggttcag
 101951 gttgctgtgg atgctcctga tgctgcgggt cgctcgtggtg tgggggtgta
 5 102001 ttcgtgtcct gatggtgtgg gtcaggcgggt gtggtcgcag catgctgtcg
 102051 gtgtgttggc ctctggtgtg gctgaccagg tcgggtgggt cggtgacggt
 102101 ggtgtgtggc cgccgcaggg tgcggtgtcg gtggatgctg agggctgcta
 102151 cgagctgttt gcggtatgctg ggttcggtta tggcccggtg ttccaggggt
 102201 tgcgtgcgggt gtggcgtcgc ggcgaggaac tcttcgcaga ggctgccctg
 10 102251 tcggacgagg ttgctgagag cgctgatacg gcgaccggtt tcgggttgca
 102301 cccggcggtt ctggatgcct cgctccatgc ctgccttctc tccctccttg
 102351 aaggtcaatc cgccgatggt gggcctgcgt tgccgttcgc gtgggaggggt
 102401 gtttccctct tcgcctcggg tgcgacggct ttgcgcgtgc ggttggcgc
 102451 ggcgggtgag catgcggtgt cggtgaccgc ggtggatccg accggtgcgc
 15 102501 cggtgatttc catcgacgcg cttcgtacct gtcgcctcac cctcgatgag
 102551 gtcaacgcat ctacacacca gctgagcgat gcgctcttcg gcgtccaatg
 102601 gaccacggtc ccgagcaccc cggccgccga ccaccgctc gtagccatca
 102651 tcggaaccga ccacctggg ctcgccgaag cgctcagcag ttctctgtct
 102701 ggcgcgacca cgaccacgac cgcggccgcg tacgagagcc ttgacgcgct
 20 102751 gatcgcggcg gggcccgaa gttccgtccc tgacgtcaca ctcatcggtc
 102801 tcaccaccga ggacgccatc gctcagtacg tgaacgacca cgacgccacg
 102851 gtggccggcc aaggcaccat cggagccggc gcggcgggcg tagacgggc
 102901 tcgccgcctc accgccgagg ccctgcgcac gatccaggca tggttggccg
 102951 acgagcgtct ggcggcaagg cgcctggtct tcgtgaccgc cggcgggcc
 25 103001 gacgggcagg atgtcgccgc ggcggccgct cagggcctgg tgctctccgc
 103051 ccagaccgag aaccgggca ccttcggcct cctcgacctc gacgggaccg
 103101 aggcgtcgac cgcggtcctc ggcgaggctc tcacctccga cgaaccgcaa
 103151 ctgcttctgc gcgacggaca cctgcacgcc gccaggctga cccgcctggc
 103201 gtcgcccgcc gacacggccg tgcccacgga gtggaacgcg gacggcacgg
 30 103251 tgctgatcac cggtggtacc ggcgggctcg gcgcgcagtt cgcacggcac
 103301 ctgctcgaca ggtacggcgt ccgcaatctc ctgctcgtca gccggcgcg
 103351 ccccgatgcc ccgggaacca cggagttggt cgccgagctg acggcgcacg
 103401 gtgccgagggt ggccgtgcag gcatgtgacg tggccgatgg cgatcggtg
 103451 gcggcgttgg tcgccggcgt gccggatgag caccgcctga gggcggtcgt
 35 103501 ccacacggcc ggtgtgttgg acgacggagt gatcggtcgc ctcaccgagg
 103551 agcggctcgc caccgtcctg cggcccaagg cggtatgccgc ctggcatctg
 103601 cacgaagcga cccgcggcct ggacctggac gcgttcgtcg tcttctcctc

103651 cgtcgcaggg gtcttcggcg gcgccggcca ggccaactac gccgcggcca
103701 acgccttcct ggatgcgctg atggcccagc gccgggcagc gggcctgccc
103751 ggactctcgt tggcctgggg gccgtgggac cagaccggcg gaatgacggg
103801 catgctgtcg gacgccgagg ccgaccgcct cgcccgcctc ggcattccgc
5 103851 cgctctccgc ggagcagggc ctgcacctct tcgacgcggc actcgctctt
103901 gccggaacca gcacgccaga cagggcagcc ggcagcgccg ccgccagcac
103951 gtcgggaacc ggcgacacga tcgccatccc ggccgcggcc ctcgctgcac
104001 cggctccggct cgacctggcg gcgctggccg cgcagggtga ggtccccgcg
104051 attctgcgcy gactggtgcy gaccgcacc cgccgtacgg cggccggcgg
10 104101 ttccgtcacc gtggccggac tcgtcaaccg cctgtccggg ctaccgcgg
104151 acgagcggcg ccaggaactt etcgaactgg tccgcactca ggcagccctc
104201 gtcctcgggc acgccgatcc ggcgtccgtg gactccaccg cacagtccg
104251 tgacctcggc ttcgactcgc tgaccgcgt cgagctgcgc aaccggctga
104301 gcacggccac cggcctgcgc ctgaccgcaa ccctggtctt cgactaccg
15 104351 aacaccgatg ccctcgcgga gcacctgcgg gacgagttgt tcggcgcggt
104401 cgagagcgag gtgcgggtgc cgggccaggc actgccgccg accgccgacg
104451 atcccatcgt ggtggtgggc atggcctgcc gtttccccgg tgggtgtgacc
104501 tcgcccaggg acctgtggcg cctggtcgac gccggcaccg acgccatcac
104551 caccttcccg accaaccgcg gctgggacct cgaatcgctc tacgaccgg
20 104601 accccgcaca cctcggcacc tcctacaccc gctccggtgg ctctctgcac
104651 gaggcggggg agttcgaccc ggcgttcttc ggaatgagcc cgcgtgaggc
104701 gctggcaacc gactcccaac agcgtctcct gctggaatcc tcctgggagg
104751 cgatcgagcg ggccggcatc gaccgcgtga ccctgcgcgg cagcgccacc
104801 ggcgtcttcg ccggcgtgat gtacagcgac tacgggagca tcctcggcgg
25 104851 caaggagttc gagggcttcc aaggccaggg aagtgcgggc agcgtggcct
104901 cgggccgctg ctctacacc ctcggttcg aaggtcccgc cgtaaccgtg
104951 gataccgcct gctctctctc cctggtcgcc ctgcacctgg cagcccaggc
105001 ccttcgggcy ggtgagtga cgtcgcgct cgccggtggt gtgacggtga
105051 tgtccacgcc aggcacgttc gtggagtctt cgcggcagcg gggctcggcg
30 105101 cctgatggtc gttccaaggc gttcgccgag gccgcggacg gcgtcggctg
105151 gtccgagggc gtcggcatcc tcgtcctgga gcgccagtcg gacgccgtgc
105201 gcaacggcca cgagatcctc gccgtgatcc gcggctcggc ggtcaaccag
105251 gacggtgctg ccaacggcct gaccgcgcc aacggcccgt cccagcagcg
105301 cgtcatccgt caggcgttgg ccagtggcgg cctgtccacg gccgacgtgg
35 105351 acgccgtcga ggcgcacggc acgggtacga cgtcgggtga cccgatcgag
105401 gccagggcg tcctggccac ctacggccgc gaccgcgacc ccgagaaccc
105451 gctgctgctc ggttcgatca agtccaacct cgccacacc caggcagcgg

105501 ccggtgtcgc cgggtgtcatc aagatggtca tggcgatgcg gcacggcgtg
105551 ctgccgcaga ccctgcatgt cgacgcgccg tcctcgcacg tcgattggag
105601 cgtcggcgcc gtcgaactgc tcaccgagca gaccgtctgg ccggagaccg
105651 gccgggtccg tcgcgccggt gtctcctcct tcggcatcag cggcaccaac
5 105701 gcccacgtca tcctggaaca gccggaggcc gtgcagcgcc tggcacggg
105751 agcagcagag accgtggagc cggtcgccat caagccgtcg gcggaaccgt
105801 ccctggtgcc gtgggcgtg tccggcaagt cacccgaggc cctgcgcgcc
105851 caggccgcac gcctccgtga cttcctggcg gaacggcccg aaccgcgctc
105901 gatcgacatc ggccactcac tggccgtcac acgctcgag ttcgaccacc
10 105951 gcgcgatcgt gctggtcgac gatgcgaagg ccccggcgga cagcctggcc
106001 gccctcgcg cctggcctc cgggtgtggc gatcccgccg tcgtctccga
106051 cgcggtatcg accggcggtt cggcagtgtt gttcacaggt cagggtgtct
106101 aacgtctggg catggggcgt gagttgtacg gccgtttccc ggtcttcgcc
106151 gaggtctctg atgtcgtggt cgaccacctg gacgcgcct tgcgcgcca
15 106201 ggccggtttg cgtgaggtga tgtggggcga cgatgtcgag ttgctgaacg
106251 agacgggttg gaccagccc gcgtcttcg ccgtcgaggt ggcgtgttc
106301 cggctggtg agcgttggg tgtcgtccg gacttcgtgg ccggtcatte
106351 catcggtgag atcgcgccg cgcatgtcgc cggggtgttc tcgctggagg
106401 acgcctgccg tctggtggc gcccggtcga cgctgatga ggcgtgccg
20 106451 accggcgcg cgatgatgc ggtccaggcc accgaggacg aagtgacccc
106501 gcacctgacc gacgaggtg cgatcgcgcc cgtcaacggc ccgacctccg
106551 tggatgctc gggcgagaa gaggccacgc agaccgtggc acaacacttc
106601 gccgaccagg ggcgccgac gaccgcgtg cgggtctcgc atgcgttcca
106651 ctcgccgtg atggatccga tgctggcgga gttccgtgcg gtggcggaag
25 106701 gactgtccta cgccacccc tccctccccg tcgtctcgaa tctgacgggc
106751 tggctggcca cggccgacga actgtgctc gccgagtact ggggtcgcca
106801 cgtccgcgag gcggtccgt tcgccgacg catcaccacc ctcgaagccg
106851 agggcgtgcg gaccttcctg gaactcggcc cggacggcat cctgtccgcg
106901 ctggctcagc agtccctcgc cggcgaagcc gtcaccgtgc ccgtcctgcg
30 106951 caaggaccgc ggtgaggagt ccacggccct gacggcccga gcgcatctcc
107001 acacccgcgg actgatcga gactggcagg acttcttcgc tgggtgtggg
107051 gcggggcggg tggagttgcc gacgtatgcg ttccagcgtg ggtggtctg
107101 gccggttggc cgggttggg ttggtggtga tgtgggtgct gtggggctg
107151 ggtctgcggg gcatccgtt ttgggtgctg cgggtggagt ggctgcgggt
35 107201 gcgggggtg tgttgacggg tcgtctgtc ttgtcgtcgc atggttgggt
107251 ggctgatcat gcggtgatg ggcgggtgtt tgctcctggt acggcgttgc
107301 tggagatggt gatgcgtgct ggtgatgagg tggggtgtg tcgtgttgag

107351 gaactgacgt tggcggcgcc gttggtgttg cctgagcgtg gtgggggtgcg
107401 ggttcaggtt gctgtggatg ctccctgatgc tgcgggtcgt cgtggtgtgg
107451 ggggtgtattc gtgtcctgat ggtgtgggtc aggcgggtgtg gtcgcagcat
107501 gctgtcggtg tgttgccctc tggtcggct gaccaggctg gtgggttcgg
5 107551 tgacgggtgt gtgtggccgc cgcagggtgc ggtgtcggtg gatgctgagg
107601 gctgctacga gctgtttgcg gatgctgggt tcggttatgg cccggtgttc
107651 caggggttgc gtgcgggtgt gcgtcgcggc gaggaactct tcgcagaggt
107701 cgccctgtcg gacgaggttg ctgagagcgc tgatacggcg accggtttcg
107751 ggttgacccc ggcgttgctg gatgcctcgc tccatgcctc gcttctctcc
10 107801 tcccttgaag gtcaatccgc cgatggtggg cctgcgttgc cgttcgcgtg
107851 ggaggggtgt tccctcttcg cctcgggtgc gacggctttg cgcgtgcgg
107901 tggcgccggc ggggtgagcat gcggtgtcgg tgaccgcggg ggatccgacc
107951 ggtgcgccgg tgatttccat cgacgcgctt cgtacccgtc gcctcacct
108001 cgatgaggtc aacgcacccc acaccagct gagcgatgcg ctcttcggcg
15 108051 tccaatggac cacgggtccg agcaccccg ccgcgacca cccgtcggta
108101 gccatcatcg gaaccgaccc cttcggcctc gcagacggc tttcggacgc
108151 cttgcccctg gtcgaggagc gcggtgacct cgcggcgctc gcagcgtcgg
108201 agcacccggg accggacctc gtcctcgtcc cggtagcggg caccggcgcc
108251 acaggcgtag ctgcggacgc cgaaggacac accgacgccg ggacatccga
20 108301 catgctccga tccgtgcgtg aggccaccgc acaggtagt gagcagatcc
108351 agcagtgtt ggcgacgac cgttcgagg cggcgcggt ggtgttcgtg
108401 acgcgcgggg cggtttccgt gggtaggggc ggcatcgccg acctggcggc
108451 ctcgccctc tggggtctgg tgcggtcggc gcagtcggag aatccgggct
108501 gcttcggtct tctcgacctc gacctcgacc tcgcccttga ctccgacctt
25 108551 gccccgagg tcgacatcga gcgcgacct gaccgcgatc cggtcggtg
108601 gaccgtgcag cccgcgctcg ccgcggccct gcacgcgacc gccgacgagc
108651 cgcagttggc actgcgcggc gggacctgc aggccgccg actgacctga
108701 atccccgcgc cgcagaccga ccgtgccgag accgacctg ccgagaccga
108751 ccgtccggag atcgacaccc ggcggcccg cacggtgtc atcaccggtg
30 108801 gtaccggtgg cctcgggtgg ttgctcgccc ggacacctgt cgccgagcgg
108851 ggggtacgga gcctggtgct cgccagccgg agcgggtctc cggccgaggg
108901 agcggagaag ctggtcgccg acctcgaagc gtcgggtgcc gtggtggccg
108951 tgcagacgtg tgatgtggcc gatggcgatg cggtgggcgc gttggtcgcc
109001 ggcgtgtcgg acgagtaccc gctgacggcg gtcgtccaca cggccggtgt
35 109051 gttggacgac ggagtgatcg gctcgctcac cgaggagcgg ctcgccaccg
109101 tcctgcggcc caaggcgat gccgcctggc atctgcacga ggcgaccgc
109151 gatctggacc tggacgcgtt cgtcgtcttc tcctccctcg ccggcgtcct

109201 cgggtggcgcc ggtcaggcca actacgcggc ggccaacacg ttcttgagcg
109251 ccttgatggc gcagcgctgc gccgccgggt tgccgggtgt gtcgctggcg
109301 tggggtccgt gggaccgggc cggcggcatg acggggaccc tgctggacgc
109351 cgaggccgac cgcctcgccc gctccggtgt tccgccgac tcggcggagc
5 109401 agggccttgc gctgtacgac gcggcgaccg ccggtgagcg gccgctggtg
109451 gtgccggtgc ggctggacct cgccgcgctc cgccggctcg gtgatgtccc
109501 ggcgctgctg cgcggactgg tccggacgcc cgcgcggcgg accgcggcgg
109551 ccggtgcggc gccgtcgccc gatgtgctca cccggcagtt ggccgggctc
109601 ggccggggcgg agcaggagga ggtcctgctg aggctggtgc gcggtcaggc
10 109651 cgcggtggtg ctcgggcacg ccgacggctc ggcgatcggg gcggggcgac
109701 agttccagga gttgggcttc gactcgctga ccgcggtgga gttccgcaac
109751 cgactcaacg cggccaccgg actgcggctg ccggccaccc tgctgttcga
109801 ctacccgacg ccggccgacg tcgtcgggca cctgcgcggc cggctcggca
109851 ccggggaggt gtcgggtgcg ggctcgggct tggcggcgct ggacaacctt
15 109901 gaggcggtga tcgccgggct gtccctcgac gacgcggggg agcaccagtt
109951 ggtggccggc cggtcgagg tctcagggc gaagtgggcg gacatgcgaa
110001 gcgcggaggg agctgtggac ggcggtgcgg acgtcgacat cgaggaggcg
110051 tcggacgacg acatgttcgc gctgctggac gacgagctgg ggctgaactg
110101 agccgctccg catgagcagt tccgcaccag aagttccaca gtggcctggt
20 110151 ccacagtgac ctggtccaca gtgagctggt ccacagcgac ctgttccgcg
110201 gcggttccgt agtgagccgt tccgcttagg cgggaaaccg gggccccggt
110251 ggtcggacgc ttgattccgg ccaccgggag cgtcaaaccg gcctcttcta
110301 aagaaaggga aagaactagg gaaacaccca cagcccgatc atgtgcaatg
110351 aattccgtgg ctcggaaaac cattcgcggg ccattggagt acggtgtgat
25 110401 cacgtgtccg tacggcagtg acggcgtgca cgggttgcca cggcaggctc
110451 ccggccgagg cgtctccccg ctgctgcccg gcggtgcgtg ccgcttggtt
110501 cgtccctagg aggtcccccc atgaccacgt ccaccgagga gagcctgtgg
110551 gcccggtgct tccatccagc accggccgcc cccgtccggc tcttctgttt
110601 cccacatgcg ggcggtcgg cctccttcta cttcccgggt tcggcccaac
30 110651 tgtcctcggg tgccgaggtg ttcgccatcc agtaccggg gcgccaggac
110701 cggcgcaagg aagccggtgt cagtacctc gcgaccttg ccgaccaggt
110751 ctacgacgag ctgcgcccc tgctgaagga gcggccgagc acgttcttcg
110801 ggcacagcat gggcgcgacg ctggccttcg aggtggcccg gcgcttcgag
110851 gccgacgacg gtgacctggt ccggtgttc gcctccggg gccgggccc
35 110901 ctcccgcgtg cgtgaaggag ccgtgcaccg gcggtccgac gacggcatcg
110951 tcgaggagct gaagctgctc gccggcacca acaccgcgct gctcggcgac
111001 gaggagatcc tgcggatgat cctgcccgcg atccgcagcg actaccaggg

111051 catcgagacc taccgctgcc cgcccgacgt caccgtccgg gcgccgctga
 111101 ccgtccctcac cggcgaccgc gacccgaaga cctccctgga cgaggccgag
 111151 gcgtggcgcg gccacaccac cggggacttc gacctcaagg tgcttcccgg
 111201 tgggcacttc ttcgtcagct ccgaggcccc ggcgatcatc gatctgtctc
 5 111251 gggcgcacct cgccggcaac ggctagcggg cgcaactgcg caggccggcg
 111301 gtgccggtct gccgcacggc ccccgcgccg cctgagacgg caccatgcc
 111351 cccggcgacg cgcgttgctg gcgcctgtgg agcaccgtcc cgcgtgcgta
 111401 cgcgggcgcg cggaggacgt gctcggccgg ccggatgtcc tggaggagac
 111451 cattcgaaa acgtcaaatg tctgacagct gctgctcatt tactccagcc
 10 111501 cgcaactatat gatcttgtgg tggggcatgt acggagcgtg ccactcgttc
 111551 gtccttggtg catccatccc gaccgtccgg atttgtgtgc accggacggc
 111601 cgtggtcctg gtacgcgtcg gcgcgaggtc ggctcctgtt ctcacccac
 111651 gtggagggaac gcgtgatgcg gaagcagtcg ggttcttccg gcttactgac
 111701 cacgtctgtg ggacgcgacg acgaactgcg caccctcgcc cggcacgccg
 15 111751 cggccgcccc cgacgggcgg gccggcctgg tcctgtctca cgggccccgc
 111801 ggcatgggca agaccagtct gctgcggtcc ttcacggcga gcgatgtctg
 111851 ccgcggcatg acggtgctgt acggcacctg cggcgagacc gtcgccggcg
 111901 ccgggtacgg cgggtgtgct gaactcctcg gcgggctcgg cctgagcggc
 111951 ggcgacgccc ggcgctcgcc cctcctggag ggccctggcg cccgcgcgct
 20 112001 gcccgcactc accgcagacc ccgccggtcc cgacgcgcgc accgggtgct
 112051 acccggtgct gcacggcctg tactggctcg ccgccgcct catggccaa
 112101 cggccgctgg tcctcgctct cgacgacgtc cactggtgcg acgaacgctc
 112151 cctggcctgg atcgacttcc tgctgcgacg ccggaggac ctgccgctgc
 112201 tggctgctct ggctggcgc agcgaggccg aaccggtcgc gcccgcggtg
 25 112251 ctcgccgaca tcgccgcca gcgcgcccc accgtgctcg gcctgcaccc
 112301 gtcggcccc gacgacatag gcgaatggt gcgtcgctc ttccggacca
 112351 cggccgcacc gtcgttcgtc agccgggtcg ccgccgtgtc cggcggcaat
 112401 ccgctggccc tcgccgcct cctcgacgaa ctccgcgcg agggcgctcg
 112451 gccggacgcc gccggggagc gccgggcccgc cgaggtcggc agtcacgtcc
 30 112501 tcgcccgtc ggtgcgctgc ctgttgagc gccggccgcc ctgggtgcgc
 112551 ggcgtggccc gtgccatgc cgtactcggc ccggagtga ccgagttgct
 112601 ggcggcgctc gccggcgctc cggccgcgac cgtcgacgag gccctgttgg
 112651 tgctgcgacg ggccggcatc ctggccgccg accgcgtgga cttcgtccat
 112701 gacgtcgtcc gtcgccggt gtcgacgac gtcgccccgc ccaccctggc
 35 112751 cgaactgcgc accaacgcc cgctgttgct gagcgacgcc ggccgcccct
 112801 ccgaggagct cgccggccag ctcatgctgc tgccggtgct cgaccagccg
 112851 tggatggccg cgggtctgcg cgacgccgcc gccagggcg agagccggcg

112901 cgccccggag gccggtgtgc gctgcctcta ccgggtgttg gaggtggagc
112951 cggacaacgt tgccgtccgc atccagatgg cccgcgcgct cgccgagatc
113001 aaccgccccg aggcgatgcg cctcctcaag gaagcgctct ccctcgccgg
113051 ggacgtccgc acccgcgccc aggtcgccgt ccagtacggc ttcacctgcc
5 113101 tcgccgtgca ggaatccccg tccggggtgc ggatgctgga ggacgcgctc
113151 gccgagctga ccgccaact gggccccgaa ccagggcccg tggaccggga
113201 gttgcggacc ctctggagt ccgtgctgct catcgtcggg gccgacgaga
113251 aggtgacgat cgggtgcggtc cgcgaccggg cggccccggt caccatgccg
113301 cccggcgaca caccggccca gcggcagatg ctggccatga ccaccgtgct
10 113351 gaccgcgatg gacggccggg acgcccggtc ggccgtcgac caggccccgc
113401 gcgccctgcg cgcacccggc gtcgagctgg aaccctggtc gctgctgtcc
113451 gcctccttcg ccctctccct ggccgacgag gtcgccgacg cgcagtacgc
113501 actggacctc atgctccagt acggccagga caacgcggcg gtgtggacgt
113551 acgtcctggc gctctccacg cgcgccctgc tccaccacgg ggtgggcgcc
15 113601 ttccccgagg ccctggccga cgcacgacc gccgtcgaga tcctcggcga
113651 ggagcgctgg gcgacggcg ccgtgctgcc ccgtgtcgcg ctggccaccg
113701 ccctcgtcga ccgcggcgag cccgagcgcg ccgaacacgt cctcgacggc
113751 atcacccgcc cccgcctcga acgcttcgtc atcgaatacc actggtacct
113801 ccaggccccg gcctacgcc gctgggtccg cggggatttc caaggagccc
20 113851 tggacctcct tctcgctgc ggtcggtccc tggaggagtc gcgcttcagc
113901 aaccggcgt tcgtgccgtg gtgggccgat ggcgcggtgc tcttgccgac
113951 cctggaccgc cagcaccagg cgcgcgaact cgcgcatac ggaagcgagt
114001 tggccgagcg ttgggggacg gcgcgcggcc tcggactggc cttcatggcc
114051 caggcgctcg ccgcacccgg ccgcgcggc atcgatcacc tcaccgaggc
25 114101 ggtctcgctg ctcgcggaact cccggccccg ggccatggag gcccgggcgg
114151 aacttcttct cggacacgcc cacctgaagc gcgacgacct gcgggcccgc
114201 cgggaacacc tgcgcgcgc cgcgcacct gccagcgct gcggcgccgt
114251 gaagctcggc gtcgacgcca gaaaactgct ggtcaccgcg ggtggtcggg
114301 tacgcaggat gaccgcctcc ccactcgaca tgetgaccg gatggaacgc
30 114351 acggtggcgg acctcgcggt gaccggcgcg agcaaccggg ccatcgcgga
114401 agccctcttc gtgactgtaa ggaccatcga aaccatctc acgagcgtct
114451 accggaagct cggggtcggc gggcgtagcg agctgtccgc cgtcctggag
114501 accaggaccg ccacctccg tcggcagccg ccggcctggg tctcccaggc
114551 acgcggacgc gcttgagacg aacaggacga gaggtagcgg tgccccggag
35 114601 caaggcgcgg aaccaaccga ccacctgcac gccgcagtgc gcacccgacg
114651 cccacggcga cccaccatg ctctcgaat gcggccggga acagcggtc
114701 atcggcgacc tcctgcaccg cctcgccag gggcgccat cgggtgctcag

114751 cctgaccggc cggcccgggc acgcccagaa cgccctggtc cgctggggcg
114801 cgtgccgggc caggcacgac gggctgcgcg tcctgcgagc ccaggcgacg
114851 cccgcggaac gggaactccg ctacggcgcc gttctccaac tgctggccgt
114901 cctcgacggc ccgcacggca gcaccctgga cgccgcgacg cgccacgacg
5 114951 gtcccccgcc actgcccggtg cccggcatcg aggaggtgct gcggcgaccc
115001 ggcacggcac ccaccctggt cgtggtcgaa gacgtccagt ggttggaacc
115051 ggcctcgctg acgtggttgc agatcctgct gcgccacctc gggccggaca
115101 ccccgctcgc cgtcctggcc agcagctgcg gtgacaccac ggccttcgac
115151 accgacccga agggccccgc cgtcccgggg ccgcccggaca ccgtgcccgt
10 115201 cgcgcgcttc gtggtgcccg cgctcaccga ccgcggggtc gccgccaccg
115251 tccgcgccgt ctgcggcacc cccggcgacg aggagttcat cgccgcgctc
115301 acctccgcca ccgcccggaa ccccgccatc ctgcgggacg ccctgcgcgc
115351 ctctgctgac cacggcctcc ccgcccgcgc cgaccacctc ccggagctgc
115401 acgccctcac cgtggtgctc gtcggcgacc acaccgtgcg cgccctggac
15 115451 ggctgccccg ccgaagtcaa cgccgtcctg cgggccctgg ccgtctgcgg
115501 cgacctgctc gaattccacc gagtccgggc cctcgccggc gcgcactcgc
115551 tgtccgagga ccgatccgc accctgctgg cgagcgctcg cctgaccgtg
115601 tccgtcgggc acaagggtgca catccgcttc cccgcctcca aggcacgcgt
115651 catcgaggac atgcccgccg cggagcggcg cgatctgtac gtccgcgcgg
20 115701 ccgaactcac ccacagttgc ggcgtcaacg acgaggacgt cgcccatctg
115751 ctgctgcgct cgtcgccgct cggcgcaccc tgggtcgtgc ccctgctccg
115801 ccgcggattc gccgcgcgcg tgcgcgggga ggaccaccac cgggcctgtg
115851 cctgcctctc ccgcgccctg caggaaaccc tcgacccccg ggaacgcagc
115901 ctgctgacgt tggaactggc cgcgggcgaa gccgtcgccc ggcggaggc
25 115951 gggggatcga cgccctggggg aactcgtccg cagcacctgc gcggacaccg
116001 accccaagtc gtccggtgag ggggtggggg tccgcgcat cgacctgggg
116051 ttcgcccggg gcaacagcga atgggtccgc cgcaccgcgg gcgaggccct
116101 gccgtacgcc gggccggccg accgggagga actggtcgcg ctgttctggt
116151 tggccgcgct gcgggacgac gacgcgccga tgatccccgt ggtgccccgg
30 116201 ttgcccgaac ggccggtgcc gccggcccag gccggcgccc gtgcctggca
116251 gctggccacg gcgggggagg acgcggaaca ggccaggaag ctgccccgga
116301 tcgcccctac cggcggggtg aacgagagcc tgatgatgcc gaaactggcg
116351 gcctgcgcgc cgtgttcgc caccgacgac aacgacgagg cgtgacacgg
116401 cctggacacc atgtcaccg ccgcccgcag tgcccacctg cgcagcatgg
35 116451 ccgcccgcgt tttaacctc cgggcccggg tacacctgtg gcggggcccg
116501 ctggaggccg ccgaacgcga tctggacagc gccgagcgcg ccctgcgcgc
116551 gacgagttgg caccgccgtg cgtgcccga cctgatcgcc acccgcatcc

116601 tcgtcagcat ggagacgggc cgcccggacc gggcccgcg actcgccgag
116651 gccccggtcc ccgccggcgg cgaggagggt gtgtggtggc ccgccctgct
116701 gctcgcccgc gcccggtgtg ccgccgacga cggtgactgg gaggaggccc
116751 tgcggctgtc gcgggagtgc gggcgctggc tttttcgccg gcactgggccc
5 116801 aaccggcca tgctcagttg gcgccgctg gccgccgagg cgtgtctgaa
116851 gctcggcgac gtgacggagg cgcgcggct gccggacgag gagctgttct
116901 tcgccgaccg ctggggcacc gcgagcgccc gcgggatcgc ccgcctgacg
116951 acgcgcgac ttttcgacga cgacggcgac cgggccgtcc ggcggatccg
117001 cgaggccgcc gccctgtcc gcgactcgcc cgcccgcctg gcctacctgt
10 117051 ggagccggct gagccaggcc ggtgccgaga cggccacgg cgacaccgcc
117101 gcggccgcac gctcctggca ggcggtcgcc cggatgaccg ccgccacccc
117151 cgccagccgc ctgccaccg ccgccgcac cctgaccgtc ccgtccgttc
117201 cggtcgccac cgcgcgccc accgccgtcg tcccaccgg atggcgcgac
117251 ctgtccgagg cggaagaagg caccgtgctg ctgcgcgcc gcggccacgg
15 117301 caaccgccag atcgccgaac aactcgccgt cagcaggcgc accgtggagc
117351 tccggctgag caacgcctac cgcaagtga ggatcggcgg acgcaaggag
117401 ctgtacctgc tcctggaggc gctggaagga ccggtcgcgg atgcttcttg
117451 agcgggagaa cgaactggcc cggatccggg ccgccctgga cgccgcggaa
117501 gcgggagact cctcgctcct cctgatcaac ggtcccctcg gcagcggcgg
20 117551 ttcggcgctg ctgcgccgga taccggagct ggccggcgac ggcacccgcg
117601 tcctgcgggc cagcgccgcc tggcggaac gcgacttccc cttcgggatc
117651 gcccgccaac ttttcgacca cctgctgtcc ggggcggcgg gcgcagggcc
117701 ggccgaacgc accgcgggg cagagcactt cagccgactg atggacaccg
117751 gcgaccgcc taccgggacc ggccccgcc tggaggcttc ccaggcagtg
25 117801 ctccagggcg cccaggcgct gctcgccgac gcgtccgcgg agcggcgct
117851 gctgacctg gtcgacgacc tccagtgggc cgacggcccc tccctgcgct
117901 ggctggccca cctcaccgg cggtgcacg gcctgcgggc gctgctggtg
117951 tgcacgctgg ccgacggcga ccaccggggc aggtaccccc tggtcggga
118001 ggtcgccggc gccgcgcaca ccgtcctgcg cctggcgccg ctgtcccggg
30 118051 acgccacccg cgtcctgtc gccgggcccc agggccggcc gccgcaggac
118101 gcactggtgc gcgccgtgta cgaggcgctc aggggcaacc cgctgttctt
118151 gaccgccttc cggagcgctc tgcgcgccac cggaaggccg cccggcgggc
118201 accacttcgg cgccgtccgg gagctgagcc cgacgggtgct gcgcgatcgg
118251 ctgcggggcc atctgcggat ccagccgag ccggtgcgcg aggtcgcggt
35 118301 ggcggtggcc gcgctggggc accacagcga tccggtgctg ctgcgccagc
118351 tcgcccgggt cgatgaaatc ggtttcgccg gtgcccgccg cgcgctggtg
118401 gacgccggcc tgttggcccc gggacgggac gtccgcttcg tccacggcgt

118451 cgtccgcgat gcggtggact ccctgctcac cctcgacgag cgggaacgct
118501 cgcacgacga cgccgccgat ctgctgtacc gctgcgggcg gccggccgag
118551 caggtcgccg gccatctgct ggccgtggtc caccggggcc ggccctggtc
118601 ggaggcggtc ctgcgtccg cgcccacaa cgcgtgcgc gccggccggc
5 118651 ccgccgacgc ggcccggtac ctgcgccgcg ccctgctgca ccaccgacc
118701 caggacggct gccgcgccg catcctggtc gatctggcca ccgccgagcg
118751 cgccctcgac ccgatgcct gtgtacgcca cgtcagccag gcggtcgcgc
118801 tgctggacac ctgcgggat cgggcccgcg ccgtgttgcg catcccgcg
118851 tccctgctcg ccgccccag ccgctccgcg gtcgagttgg tgcggcaggc
10 118901 cgccgccggg ctcgacgaac cggggcagcg ggacgaggag ggagccgacg
118951 aactcgccct ggcctggag gcgtggctgc ggactccgg ccacgagaac
119001 cccgtcgagc tggcgtcctc ggtggcgcg ctgcggcgca tgggggcacg
119051 gccgccggtg gacagcgtcg ccgaacgcga gttggtcgcc gtgctgttga
119101 gtgcgggcgc gctcagcggc cggctcagcg ccgcggagat cgccgacacc
15 119151 ggcaaccgca tcctcgaacg tgagccggcc accgccgcc atgcccacac
119201 cccgctgccg ctggtgatgc tctcgtgtt cgtcgccgag tccgtgcagg
119251 gcgtggcctc ctggctggcc agcgaacagc acaccggcg ccggtacgcg
119301 accggcgcg acgatgtgct gctgaccgcc gagcgggcct tegtccgtgt
119351 gacgcagggc cgccggccg ccgcgcggga gcacgtcgaa cgcgccctgg
20 119401 tcatggacgc cggcgactgg tcggaaccg ccgtcatgat gttcgccgcg
119451 gtcgccttcg aactgcgca cccggccttg agcgaacgca tcctggaacg
119501 gatccgcgac cgccggccg ccggactggc gtcaccgcc accggtcaga
119551 tgctccaggc cgccgtcgac gtgacttcg gccgggggcg ggacgcctg
119601 gacacgtgc tggcctgtgg ccgacgcctg gagaccgtgg gatggcgcaa
25 119651 ttccgactg ctgccctggc gtccgtatgc aatcgggctg caccagcggc
119701 tcggcgagac cgatgccgcg ctgcaactcg ccgaggacga gctgaggtgg
119751 gcccgaggat ggggtgcgac gacgaacctg gccggggccc tgcgcctgaa
119801 gggctggctg ttgcaggacg agggactgga tctgctgct gagagcgtcg
119851 agatcctccg cgcttcgtcc tacgcgacgg aactggcgcg caccctcgtc
30 119901 gtcctcgggc gtcggctgcc gggtagggcg gaggccgagg cgggtgttcg
119951 ggaggccgcg gggatcgccg ccgcctgtgg tgtccctgg ctgcgccgac
120001 gggccgaact cggctcgggc agtgccatcg tgccgcggg cgcaccctg
120051 acccccagcg agcggcgagt ggcgtcgtg gtgagccgg gcctgaccaa
120101 ccaggccatc gcgaccgaac tcggtgtgag ctccggggcg gtggagaagc
35 120151 acctcaccag cgcctatcgc aagctcggcg tctccggccg ccgcgagttg
120201 gtgaatgcgc tcccgggtcg ttgacgccg cggccgtccc gttccggacg
120251 atcacaatcc tctgtgcctg tcagccctgt gggtagggcg tcgtgtcgt

120301 acgcccgtcc ggtcatgaca gcaatcctca acaaaagctc aaacggaagc
120351 gggggcaggt gcatccaccc catcaccaat gtctcgtatg ttacttttgg
120401 cccaccgtag ttgggactct gcggaccgta cctcgtcaag tacggaccac
120451 ggatttaggg gtgcgtagat gtgtcgcgtt gacgtgccga gttgtggcta
5 120501 actacgctcc gtttcacgga aatcgaaact cttcggcgtg aatcgcgcg
120551 aacttcgatc gaattctgga ttccgtttct cagaaggcta aaagacgacg
120601 ggggttatcc gaaaggatgg cggctcgtg actatcactc acctactga
120651 ctggaatcac cgtacgaccg gcggggtgat atccgctcag accgcaccgg
120701 ccggggagtc cgtcggcccc gggctgatgg cgtccctcga ccgtgacctc
10 120751 accatcaagc acgccaacca ggagtccgc cgccgcttcg acgattccgc
120801 cggtgacgtc tgcggtcgca gttccggga cctgatgcac ccgagcgtgc
120851 agcagccgct gatgcgccag ttctccggc tgatcgaggg caagcggcac
120901 cgcttcgcct cgcacgtggc cgccgtgggc gccaggacg cggccttcgc
120951 gggcaccctc accgcctccg cggtcaccg caagaccccg gacatcgccg
15 121001 ggatcctggc gctcatggac tcgtccggcg ccgccgacgc cgccgatgcc
121051 ggcgtcgtga ccagccagaa gaagtccctg accgagatcg atgcgcgcac
121101 cctcgaaggc atcgcgccg ggctgtcgac cattccgctc gcctcgggc
121151 tctacctcag ccgccagggc gtggagtacc acgtcaccg gctgctgcg
121201 aagttgcggg tgcccaaccg cgcgcgctc gtctcgcg cctactccat
20 121251 gggcatcctg aacgtcggca cctggcccc gaaggtcgtg gacgacttca
121301 tcaagtgcg cccggccggc ctgtcccga ccggccggcc accgtcggac
121351 cgctgtcccg cagcccgtg tcgccgacgc ccgcctctca cggggcgatg
121401 cccggcggca cggcggtttc gtgccctggt cgatgtttat ccattctcac
121451 gttacgaacc cgcaatgtc atgcgcgcga ttccggccac ttttcgttcc
25 121501 gtcgtaacca aactgcccgt catatcgttt ccaactgaccg tttctgtggt
121551 ccgcggtgtg ccacctggc cccactccga cgcgcccgcg acgggtcgcg
121601 agcggcctcc gaatactcac ggtaaccccc gttcctcctg gtggagttcc
121651 ccgcgcgtca cattcctccg tgctcgtgc ggcgtggctc gcccgcttcc
121701 cccgccgcg cgacagccat tgttacaacg gcacgacaga gcgggggtgg
30 121751 tcgaccgctc cgcggtgcg cgcagtcaag cggaacctct gctcgatgtt
121801 cgtgccaccc cgaactctgt gttcttggtg accgttttca ccaatgcccc
121851 gcccgttcgg tggcgaatcc gccgccggt cctggccaca acggatcgct
121901 tcggccgccc accacggccg gcgcagccga ggcttggtgg aatcaccccc
121951 acccagtcgg ggcgcggtt tggcccggac acgcaggagg aacaccgtgg
35 122001 atgctgaggg ccgcgcagg gacatgctgg agctgatccg gcgcagcgga
122051 tccgccgatg tcgtgcgtct cgcgaggag ttcccgctca gcaaggaaac
122101 ggtccgccgg gacctgaacg tcctggaggg ccatgggctg atacggcgcc

122151 ggcacggtgg cgcctatccc atggtgcggc cgggctccga ggccgtcttc
122201 gtctcccgga ccgcacagcc gatcccgag gagtcccgga tcgccaccgc
122251 ggccgccgaa ctctcagcg aggccgagac ggtcttcac gacgagggct
122301 tcaccccgca actcatcgcc gacgccctgc cgcgcgaccg gccgtgacc
5 122351 atcgtcaccg cctcgtctcc ggtggtcagc gccttcgcga cgagcccaca
122401 ggccaacgtg ctgctcctgg gcggccgggt ccgccggggc acgacggcca
122451 ccgtcgacca ctgggccgtc cacatgctgt ccggcttcgt catcgacctg
122501 gccttcctcg gcgcggagg gatctcgcgc aggtacggcc tgaccacccc
122551 cgacccggcg gtcgccgagg tcaaggcca ggccatccgc gtcgcgcgc
10 122601 gcccggtcct cgcgggggtg cacaccaagt tcggcacggc gagcttctgc
122651 cggttcggag aggtgggcca cctggagacg atcgtcaccg gcgccggcct
122701 gcccgctgcc gagggccacc gctaccacct catgggcccc aaggttttac
122751 ggggtgtgacg ccgccggggc gtcccgcgc ccgccacgg cctggcgcgc
122801 gccggtgcgg gtcagccgcg tcgcgcgtcc gtcggttccc gccagagtgc
15 122851 gagcgtgtgc acccccaccc ggtagccctc ctcgctgagg agatggccgc
122901 cgtgcttcat gtgccagagg gcgaccgcgc gttcggtgac ggtcaggtcc
122951 gcggcccgtt cggcgagtcg gacctgaag cgctgggctg cgtcgagggga
123001 gcgcacccag gccaccaggt gcaggttggt ccggctgatg acgctcgcgc
123051 acaggcggac ctgcgcgcat ccggtgacct ggccgggtcac ctgcgcgagc
20 123101 ctgcggccg ggacctgcc ccagaagctc accgtgaccg gccactcgga
123151 cagcgggagg gcgacctgc accgggctg caacatgtcc gcggcgaaga
123201 gccgttgac gcgacggcgc acggtgtcgg ggccggcgcc gactgctcg
123251 gccagcgcgc ggtaggtggc ccggccgtcc acggacagcg cgtgaccag
123301 ctgttggtcg agctcgtcga ggacgaaggc gggggtgtcc gtgcggtgtc
25 123351 tggacgcgtc ggccggcagc cgggcgacct ggtgccggcc gagggccgc
123401 aggcgccacc gggtgccctc ggtgtgcacc gggccggcca ggtgcgtccg
123451 ggccggcccg acgcgtcca acgcggcgag gtcgtgggtc acccaacggg
123501 agagcatggc cggatcgcg gccatgacgt tgagttggag gtcacggtcc
123551 ccggtgacgt gggagagcgc caccacgtgc ggccggcgcg ccaacgccc
30 123601 cgcgacgtcg agcagtcggc cgggagcgca gtcgatctcg acgaacgcca
123651 ggcacccctg gcccgacgcg gccagcaccg gggccggatg gcagctgatc
123701 caggcggcgc cggctctgac cagccggttc cagcgcttg ccaccgtcac
123751 ggcgtccagt ccgaggacgg agccgatccg ggtccagctg gcccgggcg
123801 tgatctgcag cgcgtgcacc agggcctgat ccacatggtc cagggaccgg
35 123851 ggggtctgcc cggaatctg cgccacgggc cacctcctg cgtgttccg
123901 gcggttcgg gccgcggtc ggctcaacct tcagcctgga ctccgggtacg
123951 gccggaccgt accaggcaac ccccgagca acaggagtgg gattcggcat

124001 gtcggcagtg gaacgcgcca aggaaatgca acccgagttg gagcggctgc
 124051 gccggtcgct gcaccgggag ccggagttgg gccttgcgct gccgcgcacc
 124101 caggagaagg tgctcgcggc gctcgacggg ctgccgttgg agatcacccc
 124151 gggcaccggc ctcacctcgg tgaccgccgt cctgcgcggc ggccgccccg
 5 124201 gcggcgagtg gctgctgcgc ggcgacatgg acgcgctgcc ggtcaccgag
 124251 gagagcggtg tgccgtacgc ctcggagatc cccggccgga tgcacgcctg
 124301 cggacacgac ctgcacaccg ccggactcgt cggcgcggcc aggctgttgg
 124351 ccgaacgccg ggaacgcctg cacggcgacg tgggtgttcat gttccagccc
 124401 ggcgaggagg gccacaacgg ggccggcgcg atgatcgagg agggcgtgct
 10 124451 ggacgcggcg ggcaagccgc tcgacgcggc gtacgccctg cacgtggcgt
 124501 ccaaccaggt gcccggcggc gtggtcatca cccgctccgg caccatcacc
 124551 tcggcctccg acctgctgac ggtgacggtg cggggcgagg gcggccacgg
 124601 ctcgacgccc gccacggcca aggacccggt cccggcggcc tgcgagatgg
 124651 tcaccgcggg gcagaactgg gtgacccgcg ccttcgacat cttcgacccg
 15 124701 gtcgtggtca ccgtcggcac cttccacgcc gggacgaagg cgagtgtcat
 124751 ccccgacaac cgccgagttc caggccacgg tgcgcagtta ttccgaatcc
 124801 gcccgggacc ggctcgccga agggaaactgc cccgtctggt gcggggccatc
 124851 gcggccgccc acggcctgga agcggaggtc gactacctcc ggcactatcc
 124901 ggtcaccgtc aacgacacgg acgagaccgc cttcgcggtg gccaccgccc
 20 124951 gcgacctgtt cgcgcccggc gaggtgtggg agtcgccgat acccaccaac
 125001 ggctccgagg acttcgcgta cgtcctccgg cgcgtccccg gcgcgatgct
 125051 gctggtgggc gccgcccccg aggggagcga ctggcagcac gcgcgatga
 125101 accactcgcc gcgcgccgtc ttcgacgacc ggggtgctgta ccggcaggcg
 125151 gcgctgctga ccgaactcgc cgcgcgccgg ctcgcggcca ccgcgccggc
 25 125201 cgagcccgcc gtggcgggat gacaccgccg ccgcacgcgt ggggcccgcga
 125251 cgcgtgcggc gtacgggccc tggccggtct ccttcccggc tacttgcgcg
 125301 cggggcactc cttccacgcg aagtggtaga ggggtgtgat gctgccgtcg
 125351 gtggagtcac ggccatgtag ctggtggcct tggggtcagc cccgctgccg
 125401 t

30

SEO ID NO. 36**NysDII**

35 1 msftypvsmp wlqgreldyv teavgggwis sqgpyvrrfe eaafaayndvp
 51 fgvacssgtt altlalralg vpggdevivp eftmiasawa vtytgatpvf
 101 vdcgddlnid vsrieekitp rtkvimpvhi ygrqcdmdav lnlayeynlr

151 vvedsaeahg vrprgdiacf slfankiisa geggvclthd phlaeqmahl
 201 ramafthkds flhkklaynf rmtnmqaava laqteqltdi lalrrdiekr
 251 ydealrdipg itlmpprdvl wmydlraerr delcaylage gietrvffkp
 301 msrqpgyfsa dwpalnaarl sadgfylpth tgltaqeqef itgrirafyg
 5 351 va

SEQ ID NO. 37

NysI

10

1 mdneqklrdy lklatadlrr trrrvhkles aagepvaiig mtcrypggvr
 51 spedlwrme agehgvtpfp tdrqwdleal aaaptasggf lhdapdfdad
 101 ffgispreav amdpqgrvvl esaweaferra gidptsvkgs rtgvfigama
 151 qdyrvgpapg aegfqltgnt gsvlsgrisy tfgtvgpavt vdtacssslv
 15 201 avhlatqalr agectlalag gvtimsgpgt fiemgrqggf sadgrcrsfg
 251 dtadgtguae gvgilvlerl sdavrnghei lavvrgtavn qdgasnglta
 301 pngpsqqqvi qqalvnarla agdidvveah gtgttlgdpv eaqallatyg
 351 qnrpadrpll lgsvksnlsh tqaaagvagv ikmvmamrhg tlprrlhaee
 401 pthhvdwsqg avrlltdttd wpatgaprra avssfgisgt nahtiieqap
 20 451 epqpedaata qddaagstpa tapvvpvgvp vllsgrtpda lrgqaaalra
 501 aldtgrrpdl ldlahtlatt ragfehravl latdhpaltd gltaladadd
 551 paaapawitg ttraetrlav lftgqgaqrl gagrelaarf pafataldaa
 601 ldaftphldr plrevlwgt d aalldrtaya qpalfaveva lyrliesfgv
 651 rpdhlaghsv geivaahlag vlsladaatl vaargrlmqa lpdggamiav
 25 701 qaseadvapl laghedqvai aavngpsavv lsgaeatvta laeqaadgr
 751 ktrrlrvsha fhspmlmepml dafravvedl tlqppllpvv snltgkpatv
 801 aqltsadywv dhvrhavrfa dgidwlarhd ttaflelpgd gvlsamaqdc
 851 ldaadadavt lpalragrpe ehtlttalag lhvhgatldw tgcfaagtgar
 901 rtdlptyafq rrrywpkalq sgtadlrsvg lgaahhp1ls aavsladagg
 30 951 tlltgrlrsq thpwladhtv rgttllpgta flelavragd evgcdrveel
 1001 tlaaplllpe qggvqvqlwi gnpdvsgrrt vnvharpgdtg ddtptwahat
 1051 gvlttadasr qlpasseqgg tplagdphpa ldaaqwppag aeplpldghy
 1101 drladggfgy gpvfqglraa wrggdvvyae velpeagrsd aeafglhpal
 1151 ldaalhaapf tglgergrgg lpfswegvsl haggattlrv rltpvaddal
 35 1201 altvadgtga pvlsvdsvl rsvatqqltd aaavardalf rldwtpvqpt
 1251 atdpgpvall gadpfgl1th agfadapayp dlaalaaadg pvpttvvlsl
 1301 agtgddaadp arsahrcaae alaavqtwld hherfaaar1 vfvtrgatvg

1351 rdvaavawg lvrsaqsenn gcfaivldldp dgavgaavlv aalvsgepq
 1401 avrgdvlrva rlvrrpltev gagadgtgdg vgdgsgvsfs gegavlvttg
 1451 tgglgavlar hlvaeygvrd lllvsrsger avgagelvae lagvgarvrv
 1501 vacdvtdraa vvelvgghav savvhaagvl ddgmvgaltg erlsavlrpk
 5 1551 vdavwhlhea trglldafv vfsslagvfg spgqanyaaa nafldalmtr
 1601 rraeglppls lawgpweqsg gmtgtltdvd aerlarsgvp plsvaagglal
 1651 fdaavagtda tcvpvrldlp vlrangevpp llrslirvra rraavagsat
 1701 agnlaqrllr ldedgrdemv ldlvrgqval vlghatggdv dagrafrdlg
 1751 fdsltavelr nrlntvtglr lpatlvfdyp tvrhlatyvl dellgtdae
 10 1801 atvqpaavav addpivivgm acrypggvss pedlwrvtte gtdavsgfpt
 1851 nrgwdvesly hpdpdhpgts ytrsggflhe agefdpgffg msprealatd
 1901 sqqrllless weaieragid pvsrlgsrtg vfagvmysdy samlaspefe
 1951 gfgqsgssps lasgrvaytl glegpavtvd tacssslvam hwamqalrsg
 2001 ecglalaggv tvmtspavfv dfarqrglsp dgrckafada adgvwgsegv
 15 2051 gvlvlerqsd avrngeila vvrsgavnqd gasngltapn gpsqqrvirq
 2101 alasggltag dvdvveahgt gttlgdpiea qallatygrd reperplllg
 2151 svksnlghtq aaagvagvik mvlamrhgvv prtllhvdaps shvdwsegav
 2201 ellseqaawp etgrvrragv ssfgisgtnv hviveqapga kaiaaagaar
 2251 rtpgavpvl sgrgrsalrg qaarllghlq arpdaelvdlv alslattrsr
 20 2301 feqraavvaq drdqliaslg alaadrpdpa vvegeaagrg rtavltftggg
 2351 sqraamgrel hevqpefaaa fdavcavfdp lldrplrevv faedgsdeaa
 2401 lldetgwtqp alfavevalf rlveswgvvp dfvaghsige iaahvavgl
 2451 tledacriva aratlmqalp tggamiaiaq tedeiaahld dtvaiaavng
 2501 pqsvvisgde eaaetiaatf aergrktrkl rvshafhspr mdgmldafri
 25 2551 vaegltirap riplvsdltg rraddaevct aeywvrhvve avrfadcvt
 2601 lrdagattfl elgsdgllta maedtlgddh daelvpmra graeelaat
 2651 alarlqvrqv dvdwaaylag tgartrdlpt yafghayywp qlptpaaala
 2701 aadpadqqlw aavergdare ladilglgeq dltpldslp altswrrgnq
 2751 ekhlldtlry rvewtrlskp tapvldgtwl lvasdataad qpallldglad
 30 2801 algshgarvr rlllddscad ravlaerlar tadvdaatqv lsvlplder
 2851 addcppltrg laltvalvqa ladtgaqgrl wtatrgavst npadpvthpv
 2901 qaaawglgrg valehprlw glvdlpqvfd eragqrlagi lavkdapdge
 2951 dqvalratgv sgrrlvrhtv ealptaaeft atgtvlitgg tgglgaevar
 3001 wlaragaqhl vltsrrgpda pgaaelrael egygpsvsvv acdvadrdal
 35 3051 aavltalpee lpltgvvhta gvghygpldt lstaefaglt aaklagaahl
 3101 dalladreld fffvlfgsiag vwsgnqsay gaanayldal alhrrargla
 3151 atsvawgpwa eagmaaddav setlrrqglg lldpapamte lrravvrqdv

3201 tvtvadvdwq ryaplftsar psaliaglpe vralaadert eqdatgasev
 3251 vtrvralaep eqlrltldlv rtesatvlgh ssadavpegr afrdvqfdisl
 3301 tavelrkrig aatglslpst mvfdyptple laqylraeil gavlevagpv
 3351 atggaddepi aiigmacrfp ggvspeqlw dlvasgtdai sefpvnrgwq
 5 3401 tghlfdpdpd rpgttystqg gflheadefd ptffgispre alvmdpqqrl
 3451 llettwesfe ragirpetlr stltgtfvgs syqeyglgag dgteghmvtg
 3501 sspsvlsgrl syvfglegpa vtvdacsss lvalhlacqs lngesnlav
 3551 aggatimttp npfiafsrqr alakdgrcka fsddadgmtl aegvgvvlve
 3601 rlsdaqrngh pvlavlrqsa inqdgasngl tapngpsqqr virqalanar
 10 3651 lapgdidale ahgtgtplgd pieaqalfat ygrdrdpesa lllgsvksni
 3701 ghtqsaagia svikmvmlr hselpptlha dapsshvds agtvrlltqa
 3751 rawpetgrpr raavssfgis gtnahvileq apvadtpaee rpavapvpia
 3801 agvvpwvta rsaaalrgqa erllahaetv gtalpaagpl diglslvsar
 3851 arfehravvv ppagtdplaa lravatdgps pvvargvadp egrtvfvfpg
 15 3901 qgsqwvmgs qlldesavfa eriaecaaal aeftdwsldv vlrgvvgaps
 3951 lervdvvqpa sfavmvslaa lwrsrgvlpd avvghsqgei aaavvsgals
 4001 lrdgarvval rsqaigrala grggmmsval svdvleprlv efegrsvvaa
 4051 vngprsvvva gepealdalh arltaddira rriavdyash shqveldhee
 4101 llevlaelap rtsevpffst vtgdwldtar mdagywfrnl rgrvrfadav
 20 4151 adllaaeyra fvevsshvpl smavqeaide agvpavaagt lrrdgggtldr
 4201 flisaaevfv rgvdvdwagl fegtgasrid lptyafqheh lwavppapea
 4251 vaaadpddaa fwtavedgdv saltaalgted edsvaavlp ltswrrarrd
 4301 rstvdawryr vawkplggtl phpsltgtwl lvtadgiddt dvagaletyg
 4351 aevrrlrlde ecvdravlr rлагаedvtg ivsvlaaaer tdavpgtslv
 25 4401 lgtaltvali qalgdaeida pvwaltrgav stgradelta pvqaqvtgig
 4451 wtaalehpqr wggtdlpaa ldaraaqrla avlsgalgsd dqlairpsgv
 4501 ftrrivraea tagrpagtwr prgttlvtgg sgtlaphlar wlaqrgaehl
 4551 vlisrrgtaa pgaaelvael aesgteatva acditdrdav aalladlkad
 4601 grtvrtvvht aatielhtld attladfdrv lhakvtgaqv laellddeel
 30 4651 ddfvlysstg gmwsgsghaa yvagnaylaa laehrrangl palslswgiw
 4701 addklgrvd pqmirrsgle fmdpqlalsg lqralddnen vlavadvdwe
 4751 tyhpvytsgr ptplfdevpe vrrltaaaeq sagtvaegef aalralsda
 4801 eqqrtilletv rteaasvlg lssaedltdqr afrdvqfdisl tavglrnrla
 4851 svtgltlpst mvfdypnpaa laaylhgela garsaaagaa avptgapdad
 35 4901 dpiaivgmsc rypggvgasae dlwrialdev daisgfpadr gwdaeglydp
 4951 dpdrpgrtys vqggflrdva efdpgffgis prealsmdpq qrlilletawe
 5001 afehagidpv gqrgsrtgtf vgasyqdyas gvpnsegseg hmitgtlssv

5051 lsgrvsylfg fegpavtldt acssslvamh lacqslrnge sslalaggvs
 5101 imstpmsfvg fsrqralaed grckayadga dgmtlaegvg lvllerlsda
 5151 ranghqvlav irgsavnqdg asngltapng psqqrvirqa lansavapgd
 5201 idvleghgtg talgdpieaq allatyggdr aperplllgs vksnightqm
 5 5251 asgvasvikl vralqegvvp kslhidrpst hvdwssgaig lltertpwpe
 5301 tgrprraavs sfgisgtvnh tileqapade aptpadpprd glvpvllsgr
 5351 geaalraqaa rllafveerp eahltdlahs latsraaler raaviaadrd
 5401 tltrglrals dgrpdpglvq gtagrgrtaf lftgqgsqrp gmgrelhdry
 5451 pvfadaldev larlddgpdr plrevlfaap dsaeaalldr tgyaqpalfa
 10 5501 vevalfrllt swgltpdyla ghsvgelaaa hvagvlsldd actlvaargr
 5551 lmqalpegga mvaleaaede vlpllegltd rvsvaavngp rsvvvagvee
 5601 dvllladlfa adgrtrkrir vshafhsplm damlddfaav argltyhppt
 5651 ipfvsnvsgg lataeqvrtv dywvghvraa vrfadgidwl atqgdvhtfl
 5701 elgpdgvlsa maresltdps rtallptlrg drpeepalvt avaaahahga
 15 5751 rvdwsgyfad hgarrrtlpt yafqrerywp dttaatsaht pgsaldaefw
 5801 aaverddvaa laaslddda tvtamvpalt awrrrrgeqt eldswryrvt
 5851 wkprggatap aaltgrwlv vphdhqdrqd dataawaadv etalgtttvr
 5901 ltvtttdraa laariteaag dqgpfsgvls llplatgdag hpgapaaltl
 5951 tttavqalgd agidaplwnv trgavavgra eqvtapeqaa vwglgraval
 20 6001 elparfggtl dlpatldgga arrlravlaa tdgedavalr psgvflrrla
 6051 hapagpdtar tafdpaagtv litggtggig ghvarrlard gathllltsr
 6101 rgpaapgada lraeleelga rvtlaacdaa drdalaalla elpddaplca
 6151 vfhtagvved hvvdaltpen faavlraktv aahlheltha dldlaafvlf
 6201 sstagvlga gqgnyaana hldalaehrr shgltalsva wgpwagsgmv
 25 6251 adaaeltdrv rrggfeplap epavrallra ienddtval adidwerfqr
 6301 afaavrplpf vadlpetgra tpatatgaat glrqqlaelp eherpaavld
 6351 llrtqvaavl ghadprtved dhafrdlgfd sltilelrna lnaatglsip
 6401 atlvydlptp remadfliae llgtlptdta atvastaspk lsasfeqggt
 6451 pfddpiavig igcrfpggvt tpeelwqlld egrdgisrfp ddrqwdlaal
 30 6501 gagasdtleg gfltgvadfd arffgispre alamdpqqr llettweale
 6551 ragidpttlr gsttgfvgt ngqdyptllr rsasdvagyv atgntasvms
 6601 grlsyalgle gpavtidtac ssslvalhwa gralragecd lvvaggvsvm
 6651 aspdsfvefs tggglapdgr ckafsdaadg tawsevgil vlerlsaarr
 6701 nghqvlglir gtavnqdgas ngltapngls qqrviaqala darlrpadid
 35 6751 aieahgtgtt lgdpiearal itaygrdrda erplllgtvk snightqaaa
 6801 gaagvikmlm amrhgtlprt lhvgtppshv dwsqgtvall ddarpwprrt
 6851 qprragvsaf gvsqtnahvv veqapeteap aapaaepape atptvvpwvv

6901 sgrsrealqa qldrltahta ahparsaadv grslatdrtl fphravllag
 6951 pdgvreaara aaprtpgta flfsgggaqh almghdlyqr fpvyadaltd
 7001 vlaqfdtvld vplraalfaa pgtpeaalld qtgftqpalf avevalfrla
 7051 eswrltpdfv aghsigeiaa ahvagvfsle dactlvaara slmqqlprdg
 5 7101 amvaleated evaplltdgv alaavngprs vvvagaedav ravadrllaad
 7151 grtrrltvs hafhsplmdp mltdfarvae gltyheprip lvstllgapa
 7201 gaelrtpdyw vrhvretvrf adgvralhda gagtfveigp dgvlalttqg
 7251 tldtveagap avvvplqrrd ragdlalleg latlhthgtg pswpayfeat
 7301 gghrtldlpty afqrerywpe lgapvatapq dpaawryhet waplapeaa
 10 7351 apagralvlv pagnrdtawm tavadalga tvtaepdala eqltaagdtg
 7401 wrvvvsllaa aseglpadga wpaallatld eagvhaplwc vtrgavavag
 7451 eaptavgqaa lwglgrvaal dhpdrfggla dlpaddaha agllaahlaa
 7501 pgtaeaiavr atgvharrlv rtpaaadgat wlptgtvlv ggtggtgtmg
 7551 graarwlvre garhlvltap dgtttaadte altaelaalg aritvvdhdp
 15 7601 tapdgfaall dglpddtplt avvyapeada apgtaaela alapvtalga
 7651 altgrpldaf vlfgsiaglw gvrgraaeee sgayldafar acrdrgtpal
 7701 avawgawadl vgpslaahlr mnglpvmdad taltalsrav adgsaaeava
 7751 dvrwetfapl hheartralf dalpeargal aeaardradr ktaagdygrw
 7801 laeqpaadhd aillalvtek aatvlghadh dllepdlpfr dlghdsltav
 20 7851 dlrnqltaet gltlpatlvf dhpnpaala hiraqllege sdsaaapvaap
 7901 valgaddai vivgmacryp ggvtspedlw qlvgdevdav gdfptdrgwd
 7951 laalagdpgp rsataqggfl ydatdfdppl fgisprealv mdpqqrille
 8001 tswealerag idpatlrgsg ttgvfvvggs gdyrppaeag qwqtaqsasl
 8051 lsgrlaytfq iqqptvsvdt acssslvalh laaqalrage csialaggvt
 25 8101 vmatpvgfve fsaqqalspd grcrafsdda ngtgwsegvg mlvverlsda
 8151 rrrghrvlav lrgsainqdg asngltapsg paqqrvirqa lanarlrpad
 8201 idaveahgtg trlgdpieaq allatyggdr erpvllgslk snightqaas
 8251 gvggvikmvl amqhgelpsr lyaenpsshv dwtagrahll tartpwpdsg
 8301 rprraavssf gasgtnahai leqppreelp arpaddgapl pfillsgrsqn
 30 8351 alraqarrll arltahpdtr aadlaylat traafehrea itatdhgdlr
 8401 tgltavaegt taphtaehhl qgtgkravlf sgqgsqrlgm grelherhpv
 8451 faeafdsvla rlddrltdpl rdvvwtdee alhatgntqp alfavevaly
 8501 rlieswgvvp dfvaghsve laaahvagvl slddacrlva araalmqrlp
 8551 aggamiavea tedevtpllt dgvsaaavng ptavvlsgag davtalggal
 35 8601 aerghrttrl rvshafhshl mdpmladfrt vaegleyhpp ripvvsnlgt
 8651 dvadaadlcs adyvvrvhrg tvrfadgvrt madrgvhlfl elgpdavlsa
 8701 marqcapdav vvpalrrnrd edetlvgava rlvhvgagpr wdayfagrga

8751 qwldlptypf qrgrfwpesl pgaasaapaa gqpaetdaaf wdavaqedft
 8801 alesvldves dalskvlpal mdwrsrqade sqlagwrhri vwkrltgaal
 8851 ahrkalsgtw lavvpegfad dpwvtttldg lgthlvhlev aeadraalad
 8901 aiaartadgt rfggvislla lreeltgavp egtaltttll qalgdagvda
 5 8951 plwcvtrsav sagrtldrphr plqgavwglg rvaaleypqr wggldvlpee
 9001 pdersaagla avlagldged qvavrgtavl arrlvpapgr kpsrpwhpsg
 9051 tvlvtggtga lgahvarrla kdgaqhlvll srrgpdapga aelraeldal
 9101 gtdvtvaacd vadrdqltav ldalpadrpl tgvvhtagvl ddgvldrltp
 9151 erfqevfrak vtsallldel trdrelaafv lfssasaavg npgqanyaaa
 10 9201 navldalaeq rrvlglpats vswgawggg madadgadea arragvgamd
 9251 phlaveallr lvaekptav vaevaldrfa gafggsrpsa llrefpgyre
 9301 alaaqaeqaa dggglaarla alpparrldt vvdldrtraa qvlgydpdea
 9351 vaaersfrdl gvdslgavel rnqlsaatgl nlpatlvfdh ptplvlgehi
 9401 lggflpdepa gsddeteira llasvpldql reigvlepll qlagrgraa
 15 9451 dgddgesvds mtvadlvraa lngqsdl

SEO ID NO. 38**NysJ**

20 1 mnapenpetp ennvvaalra avketdrllr qnrmlvaaak epiavvgmac
 51 rfpgevdspe alwemvatgt dvisgfpddr gwdlealrns gtdardtdvs
 101 qrggfldcia dfdpffgis preavtmdpq qrlilttawe averagidat
 151 tlratrtgaf igtngqdyay llvrslddat gdvgtgiaas aasgrlsytl
 25 201 glegpaltvd tacssslval hlavqalrng ecgmalaggv nvmatpgslv
 251 efsrqgglar dgrckafada adgtgwsega gvlllerlsd aqrnghpvla
 301 vvrqsavnqd gasngftapn gpsqqrvirq alanaglatg didaveahgt
 351 gtplgdpiea qsilatygqd rahpvllgsi ksnmghtqaa sgvagvikmi
 401 mamrhgvpr tlhvdrpsth vdwttgsvel ltdahpwpet grprrtgiss
 30 451 fgvsqtnahv iveqapdtpa eaaddtpprt prtllpwlisa rtgaalrdqa
 501 talldhldrp dgdrqptald tafslattralehrlavvt gtdgtagrda
 551 ltawlahgta pdaheghaag rtrcaalfsg qgaqrlgmgr elharfpvfa
 601 raldtavdll daelggltre viwgtddapl netgftqpall favevalyrl
 651 ieswgvapdf vaghsigeia aahvagvfl edactlvaa aglmqalprg
 35 701 gamvaveate devsplltdg vaiaaingpt slvsgdeta tlavaarlae
 751 qgrttrllrv shafhsplmd pmlaefrava eglsygepqi pvvsnltagv
 801 adgtllgtad ywvrhvrea rfadgiralt dagvgaflel gpdgtlaala

851 qqsapdavsv pvlrkdrdee paavaalarl htagvpvdtw afyagtqahr
 901 tdlptyafqy erywpkatyr padatglglt aadhpllgaa msvagsdell
 951 ltgtlslath pwladhvvvg mvffpgtgfl elavraadqv gcdrveelml
 1001 aaplilpatg tvqmqiavga adddggrdlr fftrpgddpd aawaqhatgr
 5 1051 itegervlal dtttwpprda epvdidglyd ryrangldyg pvfrglravw
 1101 rrdteiyaev alpegtadad afglhpald avlhstlfas adgddrsllp
 1151 fawngvslha agadalrvri tscgpdavei tavdpqgrpv vsvesltlra
 1201 agpdagtadh radagslfrm dwtprtvhap atpatwavlq tdpigl teal
 1251 taagpdtvtg lrdgvdalge ltagddrppv dvvavplrga tdhgpagahd
 10 1301 ltrtvalllq ewlaeerfar srlllvtrga vadgergpld laaapvwglv
 1351 rsaqsenpgr lllvldddta esaaqlpllp alldadepqa vvregtvrvg
 1401 rlarldsgrg lvpppgtpwr lgsrakgsld glallphpea rrpltghev
 1451 vgiraaglnf rdvlnalmy pgdaglfgse aagvvvevgp evtglapgdr
 1501 vmgmlfggfg plgiadarll tpvpadwswe tgasvplvfl tayyalkelg
 15 1551 glragekvlv hagaggvgma aiqiarhvga evfatasegk wdvrlslgva
 1601 ddhiassrtl dfeaaafaeva gdrgladvln alsgefvdas mrlldggrf
 1651 lemgtkdira adsvpdglly hsfldgmvd ehiqrmlldl velfdr gala
 1701 alpvrswdvr rageafrfms laqhigkivl tvppqldpdg tvlltggtg
 1751 lagllarhlv tehgarhlly agrrgpdapg aaalhaelta lgaevtvaac
 20 1801 dvadrtalaa llatvpaehp ltavvhtagv lddgtltaln pdrlatvlrp
 1851 kvdaawhlhd ltrhldlaaf vlysstagvm gggpgqanyaa gntfldalaa
 1901 hrhalglpat slawgawegg agmtgalt dh dlrrvsdagq qpllaergl
 1951 alydaataad eplivplglt ggalpagvgv pavlrglvrt agrraragta
 2001 gvsvraglaer laalpeeert pflvelvrte aatvlghgst dpvdarre fr
 25 2051 qlgfdsltai elrnrlgkat gltlpatlif dytpdrlav hlhdellgad
 2101 apvtvtaaaq aadpehdvv ivgmscrfgp gvsspeelwd lvasgtdait
 2151 gfpadrawdr hpqlagapga rtgqggflrd iadfdaffg isprealamd
 2201 pqqrilleva weaaeragid pqtllrgsdg vfmgvsgqdy aglvmsrdd
 2251 iaghattgla vsvvsgrlay alglegpals vdtacssslv slhlaaqalr
 30 2301 agectmalag gvtvmtaan ftgfsrmggl aqdgrckafs dsadgtgwse
 2351 gaavlvlerl sdarraghrv lavvrgsavn qdgasnglta pngpaqqrvi
 2401 rgalanaglt pvdvdaveah gtgtplgdp eaqaliaayg tdrdpehpl
 2451 lgsvksnigh tsaagaagl vkmvmamrhg ilpqtllhte psshdwsag
 2501 tvrllterta wprtdrprra gvssfgisgt nahvileqpp aeptpaadpg
 35 2551 rpaptvawp vsaqtpaald aqldrlrta alapltdaht latgrslfeh
 2601 ravllatvgd patgapdlpe vargaatphr tafllsgqga qrsgmgrelh
 2651 aafpvfaaaf devvavldae lgsdadggvs lrevmwgggs elldtrftq

2701 palfaveval frlvaswgv pefvaghsvg eiaaahvagv fslvdacrlv
 2751 varaslmdal pvggvmvave aaeaevvpll vdgvaiiaavn gpvsvvvsgv
 2801 eaavgqvvdq lvergrrvrr lavshafhsp lmdpmlafr avaegleyhq
 2851 pripvvsnt gevaaeeelc aadywvrhvr atvrfadgvr tlaergataf
 5 2901 leigpdgvls alargvlpae alvtptlrkd rdeesallag larlhvagvt
 2951 vdwsaaltgt gargtdlpty afqrerywpe laaepagga daadaefwaa
 3001 veradataa ahldidgdql gavlpalsaw rtrrrttsat nalrhreswe
 3051 plslagtpht ggvlvlpaa attdpwwadv vaalgpdarr vdvpadgtdr
 3101 aalaalltea addtaptavv slldaldetsg ddavpagtta taalvqalad
 10 3151 tgapaplwal trgavaalpd eqptapaqaa vwglgriaal elprhwgglv
 3201 dlpaddert arrlpaalad agdedqlalr atgaygrrit papapddapg
 3251 tgwqptgtvl itggtgalgr htarwlaahg aehlillsrs gpdapgaael
 3301 tteltalgar vtlvacdaad reqltrvlae vprdcpltg vhtagvlddg
 3351 vltgltpdrf atvfrakvas avlldeltrd tdlavfalfs svagavngp
 15 3401 qagyaanav ldalaarra qglagtsiaw gawagdgmaw rhtrpgaepv
 3451 glldpdlavp alaravtepq ptlvldlqq prllesllal rpsllsrlp
 3501 aartaaravq eadrrragaa adlrdqlagt apadrhavl rlvrttaav
 3551 lghtgadair adkpfrdlgf dsltavelss alaaatglal ppslvfdhps
 3601 praladhlra eltgdrpasa paappavpa adddpivvvg macrfpggt
 20 3651 tpeefwqla egrdgidafp tdrdwldvl grrrpgpqr prsaassyda
 3701 aafdpgfddi sprealamdp qrrlletaw eavertgtdp trlrgsrtgv
 3751 fvgtnqqdya glvlraqedv eghagtglaa svisgrlaya fgfegpavtv
 3801 dtacssslva lhwavqalra gecslalagg vtvmsttsf agftrqggla
 3851 pdghckafsd sadgtgwsseg vgvlvverrs dalrngheil avvrgsavng
 25 3901 dgasngltap ngpaqqrvir qalanaglap gdvdaveahg tgtvlgdpe
 3951 aqallatygg drpadrplwl gsvksnight qaaagaaglm kmvlalqhg
 4001 lprtlhvtep strvdwsaga vrlttertvw prtdrprrag vssfgisqtn
 4051 ahvileqppa eptptapadr ptrtpavlpw vvsarsatal daqlarlraf
 4101 aaerpdppa dvahslvtsr atfehravll aapdgitaaa raearersta
 30 4151 flfsgggaqr sgmgrrelhaa fpvfaafde vvavldaella tsggggvsrl
 4201 evmwgggsel ldrtrftqpa lfavevalfr lvaswgvgei fvaghsvgei
 4251 aaayvagvfs lvdacrlvva raslmdalpv ggvmvaveaa eaevvpllvd
 4301 gvaiiaavngp vsvvvsqvea avgqvvdqlv ergrrvrlla vshafhsplm
 4351 dpmlafrav aegleyhqpr ipvvsntge vaaaeelcaa dywvrhvrat
 35 4401 vrfadgvrtl aergatafle igpdgvlsal aaacldtda evvpalrkgr
 4451 peehtaltaa aqlhvagvdi dwtavlagtg grrialptya fqrerywpsl
 4501 aaqapgdagg lgleagrhl lgaattvags aeilltgrls ttaqpwlavv

4551 eadgrtvlpa avlaelavra gdqadcptva eltvaaplvl tgaaaqrlqv
 4601 rvaapddtgr ralsvharpd dspdspwtlh atavlthdtp qppapdtgwp
 4651 peravpldal ptatgparia aawqwgdelc aeielpepgp aerafalhpa
 4701 lldtavragg lldgdatlda lgwrglalha asatalrvrl tpdgtdtwal
 5 4751 eatdpqgapv vsvtgltlgt ptvdrsgaga addgatlldl ewvpapqaap
 4801 tggdhlpyav lgdqlaeldg qlriagdpgp rvaslaalld ggaplprlvi
 4851 apvlgvptge gdlpaavrgt ttavlellqr wtadartads hlviivtrgav
 4901 aagaedvhd1 aaapvwglvr saqsehpgsf lllldldpadp agasraaapa
 4951 tlaalldage tqaavrattl tvarltraad gpeataghpv rdwdrdgtvl
 10 5001 itggtgglgg llarhlvtgh gikhlllagr rgpdapgara lrdelaalga
 5051 evtvaacdva draaldrlla qlppehplta vvhtagvldd atvgtltper
 5101 ldtvlrakad aawhlhdatr drdlagfvly ssvagvtggp gqgnyaagnt
 5151 fldalaahra aqglpglsia wgpwgqdagm tgtlgaadla rlersgmppl
 5201 tpeqglalfd aagargdgfa vavrlargaa apgadevpav lralvrgrrr
 15 5251 taaaaghagv larllaalda eqrhqalldl vrtetaavlg hsgadavpae
 5301 rdfnrlgfds lmavelrtrl atatgarlpa tlvfdhptpd avarhlstl
 5351 pggtaagpdr splaeldria aelspegadd atrqgvvgrl rhllaqwdgt
 5401 rqdgggttvd drieaasae vlafidhelg rqads

20 SEQ ID NO. 39

NysK

1 mpdekkldvy lkwtkdilhq trqrlqevea grhepvaivg macrfpggvr
 25 51 spedlwells agrdgigpfp adrgwdlaal agdgpgrsat qeggflpdaa
 101 afdpgffdis prealamdpq qrlilletawe aversgidpa glrgsrtgvf
 151 vgtngqdyah lvlaaqddmg gyagnlaas vlsgrlafal glegpavtld
 201 tacssslvtl hlaaqavrag ecglalaggv tvmttsssf a gfsllqgglap
 251 dgrckafaea adgtgwsegi glllverlsd aqrnghpvla vlrgsavnqd
 30 301 gasnglsapn gpsqqrvirq alagaglvpg dvdaveahgt gtrlgdpiea
 351 gallatyggd rpadrplwlg svksnlghtq aaagvagvik mvlalrhgvl
 401 pqtllhdaps shvdwesgav rlltapvaws egddrvrrag vssfgisgtn
 451 ahvileqapd qpeptaeeta aaapggtaee raaapvapra vpwvpaarta
 501 galdaqlvr v ralttapgrt aadvghalat artpfehral lvheggavte
 35 551 vargavptgd rgglavlfsg qgsqrpgmgr elharypvfa aafdetvall
 601 darlgtslrd ivwdqdrtrl ddtrhtqpai favevalyrl laswgirpdh
 651 vtghsigeit aahvagvltl adactlvaar atamselppg gamvaleate

701 devrplltdd laiaavnapr svvvagaeda alavrrhfd d lgrtrtlrpv
 751 shafhsplmd pmldafrtal apltfaepi pvvsnlgtlp ataeelatph
 801 ywvchvrqav rfgdgvrata drgvrtflel gpdgvlsalv renlpepglv
 851 avpvlrkerp eettvlaalg tlwahgadvd wdavfagtrt pqadpvelpt
 5 901 yafqrarywp tlgarhgdp d lgtqtaahp llgaavtlad adetvltgrl
 951 alpshpwlgd hrsdgritvp gvafaelavr agdlsgtphl arldlpaplt
 1001 lgdgdvtltq vrvgapdpag hrpltvharl aatedapwtt catgllapda
 1051 peapadpigg adagwpprda rpvpvadlda aataagrhyg phfqgltglw
 1101 rrdgevfav alptataadr afgihpalla talrataald ddhtaghtpe
 10 1151 ptgitglalh atgatalrvr ltatgpdtda laaadatgga vltadtvtlg
 1201 spqdrpapap aghtgggglf hldwvpvdp g sratgtrwav vgddeldlgy
 1251 alhradetvs ayaaslggai gdsqglapdvf lvpvvggpa gpdavhavta
 1301 ralglqewl neprlagarl vfvtrgavav pgetvtdpag aavwgllrsa
 1351 qtenpgslll vlddafrsa gmlphvltld eqqlvvrdha vraarlarlp
 15 1401 epaagtapar awdpdgtvli tgggtgglgaa larhlvtvrg arhlillagr
 1451 gpeapgagel vaeltaqgad vrvaacdvgd rtaldallat vpaahpltav
 1501 vhtagvllda ligsitpdql atvlrpkada awhlhdatrg ldlagfvlys
 1551 svsgvlgspp qgnyaaanay ldalarhrad qglpalslaw gpwgrgsgmt
 1601 asvsdadler margglpplt vedglalfda avgrpepalv psrinvaglr
 20 1651 dqgalpalwr dlvrarrta atadrspvtv rerlrhldet gqeqlldlv
 1701 vgytagllgh pdptavdper gflelgfddl vsvglrnqla eilglrlpss
 1751 ivfidskspvk larwlhqela ngpqp gatgp aaadarpavr sddtleglfy
 1801 navrggklve amrmlkavan trpmfdtpae leelsepvtl adgpgprpli
 1851 fvsapgatgg vhqyariaah frgsrhvsal plmgfapgel lpatseaaar
 25 1901 ivaesvmas egepfvmvgh stggslayla agvledtdwv rpeavvldt
 1951 asirynpgeg ndldrttrfy ladidspsvt lnsarmsama hwfamtdiq
 2001 apaptaptll vraaraldgf rldtssvpad evrddidahl slakehsalt
 2051 aqaiiegwlae lpdpa

30 SEQ ID NO. 40

NysL

1 mstptappsl kaevppvlrl spllrelqsr apvckvrtpa gdegwlvtrh
 35 51 telkqlhdd rlarahadpa napryvhnfp ldllvddfd lartlhaemr
 101 slftpqfsar rvmldtprve alaegvlahf vaqgppadlh ndfslpfsls
 151 vlcaligvpa eeqqkiaal tklgelddpa rvqegqdel gllsglarrk

201 ritpeddvis rlclkvpsde rigpiasgll fagldsvash idlgtvlfiq
 251 hpdqlaaala deklmrgave eilrsakagg svlpriyatad vpigdvтира
 301 gdlvllldftl vnfdrtvfde pelfdirrap nphltfghgm whcigaplar
 351 vnrlrtaytll ftrlpglrlv rpveelrvls gqlsagltel pvtw

5

SEO ID NO. 41

NysM

10 1 vritvdpgrc vgagqcvlta pdlfdqdddg lvtvlagaad aadpgdvrd
 51 aalcpsgais vaad

SEO ID NO. 42

15 **NysN**

1 msteadarta apqcpvafpl rrpgrpfppp eyatyrggag lvrselfpsgp
 51 vwlvrthedv ravltdpris adpskpgfpk agrtggapsq yevpgwfvam
 101 dppehgrfrk tlipeftvrk vrelrpviqq ivderidaml aagtsadlve
 20 151 sfalpvpplv issllgvpkv drdffedrtr vlvrilstde erdkatqall
 201 rylgrliqik qrrpgddlis rliaagtlr qelsgvamll liaghettan
 251 niglgvvqll tnprwigddr iveellryys vadlvafrva vedveiggql
 301 iragegivpl iaaanhdata faapsefdpe rsarshvafg ygvhqclgqn
 351 lvreemdiay rtlfaripsl tlavpveelp lkydgvlfgl helpvtwk

25

SEO ID No. 43:

Nys RIV (long) 266 aa

30 VTITHLTDNSHRTTGGVISAQTAPAGESVGPGLMASLDRDLTIKHANQEFRRRFD
 DSAGDVCGRSFRDLMHPSVQQPLMRQFSRLIEGKRHRFASHVVAVGAQDAAFAGT
 LTASAVTGKTPDIAGILVLMDSSGAADAAGVVTSSQKKFLTEIDARILEGIAAG
 LSTIPLASRLYLSRQGVYHVTGLLRKLRVPNRAALVSRAYSMGILNVGTWPPKV
 VDDFIK

35

SEO ID NO. 44:

GCC GGC ATG CGA CGA ACA GGA CGA GAG GT

SEQ ID NO. 45:

GCC GTG GTC GAC GAA GGC

SEQ ID NO. 46:

5 CTC AGC ATG CCG AAA GGA TGG CGG

SEQ ID NO. 47:

AGG CAA GCT TCG GCG ACA CGG GCG T

SEQ ID NO. 48:

.10 CTC AGC ATG CGT ACG ACC GGC GGG

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/59126 A3

(51) International Patent Classification⁷: **C12N 15/52**,
C12P 17/18, 19/62, C07H 17/08

ALPHARMA AS [NO/NO]; Harbitzalléen 3, N-0212
Oslo (NO). SINVENT AS [NO/NO]; Strindveien 4,
N-7465 Trondheim (NO).

(21) International Application Number: PCT/GB01/00509

(22) International Filing Date: 8 February 2001 (08.02.2001)

(71) Applicant (for GB only): **DZIEGLEWSKA, Hanna**
[GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria
Street, London EC4V 4EL (GB).

(25) Filing Language: English

(71) Applicants and

(26) Publication Language: English

(72) Inventors: **ZOTCHIEV, Sergey Borisovich** [RU/NO];
Elgester gate 51, N-7030 Trondheim (NO). **SEKUROVA,**
Olga Nikolayivna [RU/NO]; Elgester gate 51, N-7030
Trondheim (NO). **FJÆRVIK, Espen** [NO/NO]; Lianveien
10, N-7025 Trondheim (NO). **BRAUTASET, Trygve**
[NO/NO]; Martin Stokkens vei 74, N-7027 Trondheim
(NO). **STRØM, Arne, Reidar** [NO/NO]; Konvallvegen
38, N-7050 Trondheim (NO). **VALLA, Svein** [NO/NO];
Naustanveien 2, N-7560 Vikhamar (NO).

(30) Priority Data:
0002840.7 8 February 2000 (08.02.2000) GB
0008786.6 10 April 2000 (10.04.2000) GB
0009387.2 14 April 2000 (14.04.2000) GB

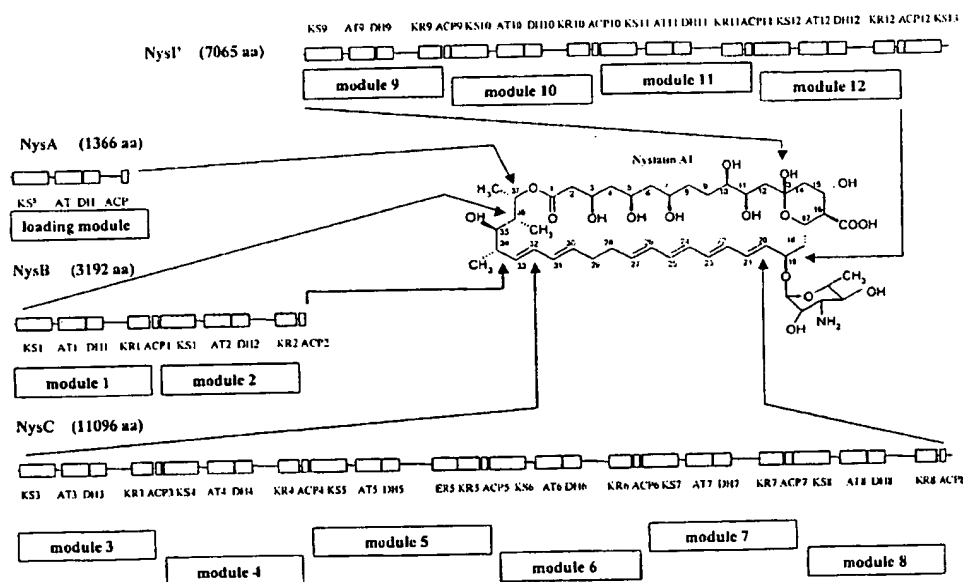
(71) Applicants (for all designated States except US):
NORGES TEKNISK NATURVITENSKAPELIGE
UNIVERSITET [NO/NO]; N-7491 Trondheim
(NO). **STIFTELSEN FOR INDUSTRIELL OG**
TEKNISK FORSKNING VED NORGES TEKNISKE
HØGSKOLE [NO/NO]; Sintef Applied Chemistry Di-
vision, Sem Seelands Vel 2A, N-7034 Trondheim (NO).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ELLINGSEN,**
Trond, Erling [NO/NO]; Markaplassen 273, N-7054
Ranheim (NO). **SLETTA, Håvard** [NO/NO]; Uglaveien

[Continued on next page]

(54) Title: GENE CLUSTER ENCODING A NYSTATIN POLYKETIDE SYNTHASE AND ITS MANIPULATION AND UTIL-
ITY



(57) Abstract: The invention provides a nucleic acid molecule comprising: a nucleotide sequence as shown in SEQ ID No. 35; wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof. Also provided are part of such molecules and polypeptides (and parts thereof) encoded by such a nucleic acid molecule, and the use of such molecules and polypeptides in facilitating nystatin biosynthesis and in the synthesis of nystatin derivatives and novel polyketide as macrolide structures.



19, N-7024 Trondheim (NO). GULLIKSEN, Ole-Martin [NO/NO]; Grevlingvn. 31 D, N-0595 Oslo (NO).

(74) Agents: DZIEGLEWSKA, Hanna et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

... with international search report

(88) Date of publication of the international search report:

14 March 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 01/00509

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 C12N15/52 C12P17/18 C12P19/62 C07H17/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL, WPI Data, PAJ, BIOSIS, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EM_PRO 'Online! EMBL; ID SNA132222, AC AJ132222, 20 April 1999 (1999-04-20) APARICIO J F: "Streptomyces natalensis pimS1 gene" XP002174375 Note: 60.5%, 60.4% and 60.5% nt seq identity with SEQ ID NO:35, 1 and 2 in 6853, 5454 and 6853 nt overlap (5017-11675:6892-13554, 5974-11277:29502-34804, and 5017-11675:6892-13554), respectively page 4-5	1-7,9, 14,17, 19-24, 27,30-36
X	-& APARICIO J F ET AL.: "The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimarin: A new polyketide synthase organization encoded by two subclusters separated by -/--	1-7,9, 14,17, 19-24, 27,30-36

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

8 August 2001

Date of mailing of the international search report

22/08/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040. Tx. 31 651 epo nl.
 Fax: (+31-70) 340-3016

Authorized officer

van de Kamp, M

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 01/00509

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	functionalization genes." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 274, no. 15, 9 April 1999 (1999-04-09), pages 10133-10139, XP002174372 abstract page 10139, left-hand column, line 12-14 ---	
X	DATABASE EM_PRO 'Online! EMBL; ID SCI7, AC AL096743, 1 July 1999 (1999-07-01) SEEGER K ET AL.: "Streptomyces coelicolor cosmid I7" XP002174376 Note: 68.3% nt seq identity (71.3% ungapped) with SEQ ID NO:35 and 1 in 1077 nt overlap (13366-14424:121753-122803 and 13366-14424:61492-62542, respectively) page 6, line 70 -page 7, line 18 page 16 ---	1-6,14, 15
X	DATABASE SWALL 'Online! SWISSPROT; ID Q9X9X5, AC Q9X9X5, 1 November 1999 (1999-11-01) SEEGER K ET AL.: "PUTATIVE TRANSCRIPTIONAL REGULATOR" XP002174377 Note: 65.6% aa seq identity with SEQ ID NO:13 in 250 aa overlap (9-258:4-253) the whole document ---	1-6,14, 15
X	SHENIN Y D ET AL.: "Nystatin: Methods for preparation, search for derivatives, and prospects for application in medicine (review)." KHIMIKO-FARMATSEVTICHESKII ZHURNAL, vol. 27, no. 2, 1993, pages 14-21, XP000984316 ISSN: 0023-1134 the whole document figure on page 17 ---	35,36
X	US 5 981 721 A (MOHAN ARTHUR G) 9 November 1999 (1999-11-09) the whole document example 12 --- -/--	35,36

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 01/00509

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MCDANIEL R ET AL: "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 96, 1999, pages 1846-1851, XP000910246 ISSN: 0027-8424 abstract; figure 2</p> <p>---</p>	32, 33
Y	<p>MALPARTIDA F ET AL: "Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes" NATURE (LONDON), vol. 325, no. 6107, 1987, pages 818-821, XP002075972 ISSN: 0028-0836 the whole document</p> <p>---</p>	1-28, 30-36
Y	<p>DECKER H ET AL.: "A general approach for cloning and characterizing dNDP-glucose dehydratase genes from actinomycetes." FEMS MICROBIOLOGY LETTERS, vol. 141, no. 2-3, 1996, pages 195-201, XP001010562 ISSN: 0378-1097 the whole document</p> <p>---</p>	1-28, 30-36
Y	<p>KHOSLA C ET AL: "Generation of polyketide libraries via combinatorial biosynthesis" TRENDS IN BIOTECHNOLOGY, vol. 14, no. 9, 1 September 1996 (1996-09-01), pages 335-341, XP004035738 ISSN: 0167-7799 the whole document</p> <p>---</p>	23-28, 30-36
A	<p>LEONARD KATZ: "Manipulation of modular polyketide synthases" CHEMICAL REVIEWS, vol. 97, no. 7, 1997, pages 2557-2575, XP002103748 ISSN: 0009-2665 paragraphs I and VII</p> <p>---</p> <p>--- -/--</p>	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 01/00509

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SEKUROVA O ET AL: "Molecular cloning and analysis of a pleiotropic regulatory gene locus from the nystatin producer <i>Streptomyces noursei</i> ATCC11455." FEMS MICROBIOLOGY LETTERS, vol. 177, no. 2, 15 August 1999 (1999-08-15), pages 297-304, XP001015583 ISSN: 0378-1097 cited in the application abstract ---	
A	VANDEN BOSSCHE H ET AL: "Anti-Candida drugs - the biochemical basis for their activity" CRITICAL REVIEWS IN MICROBIOLOGY, vol. 15, no. 1, 1987, pages 57-72, XP001015427 ISSN: 0045-6454 paragraphs I and II ---	
P, X	BRAUTASET T ET AL.: "Biosynthesis of the polyene antifungal antibiotic nystatin in <i>Streptomyces noursei</i> ATCC 11455: Analysis of the gene cluster and deduction of the biosynthetic pathway." CHEMISTRY & BIOLOGY, vol. 7, no. 6, June 2000 (2000-06), pages 395-403, XP000953274 ISSN: 1074-5521 the whole document ---	1-28, 34-36
P, X	WO 00 77222 A (COLINA ANGEL JOSE ; MARTIN JUAN FRANCISCO (ES); APARICIO JESUS F (E) 21 December 2000 (2000-12-21) the whole document page 18, line 5-17 examples 1-4 figures 3,4 ---	1-7,9, 14,17, 19-24, 27,34-36
T	ZOTCHEV S ET AL.: "Identification of a gene cluster for antibacterial polyketide-derived antibiotic biosynthesis in the nystatin producer <i>Streptomyces noursei</i> ATCC 11455." MICROBIOLOGY, vol. 146, no. 3, March 2000 (2000-03), pages 611-619, XP001012427 ISSN: 1350-0872 the whole document -----	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 01/00509

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5981721	A	09-11-1999	NONE	
WO 0077222	A	21-12-2000	AU 5407400 A	02-01-2001